Validation Studies on a Screening Tool for Mental Disorder of Children and Youth in Canada and the Implication for Program Evaluation

by

Xuejing Jiang

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences

University of Manitoba

Winnipeg

Table of Contents

List of Tables

List of Figures

Abstract

Background: The early detection and treatment of children with mental disorders hold the best promise for maximizing population health outcomes. However, child welfare resources are often limited to conduct in-depth assessments of mental wellness. Therefore, we studied the Strengths and Difficulties Questionnaire (SDQ) and its scoring metrics for the potentials in quick and convenient mental-disorder screening for Canadian children.

Study purpose: The following research questions were addressed: How well do the SDQ symptom scores diagnose mental disorders among children? Are the SDQ cut-offs for classifying as normal and abnormal proposed in other countries applicable to Canada? Does combining the SDQ symptom scores and the SDQ impact supplement yield additional predictive value? Does it have different implications for the program evaluation whether using the SDQ symptom scores as continuous variables or categorizing the symptom scores into categories (normal vs. abnormal) as categorical variables?

Study Designs and Sample: The study was conducted using linked administrative health databases and PAX program data. PAX is a classroom-based early mental health prevention program. A population-level Randomized Controlled Trial (RCT) pilot study was conducted in 2011/2012 in schools across Manitoba. This thesis uses three data sources: 1) PAX RCT pre- and post-PAX mental-health outcome measures collected in 2011/2012; 2) Manitoba Grade 5 Mental Health Survey in 2015/2016; and 3) longitudinal health, education and social services administrative data up to the fiscal year 2016/2017 held in the Population Health Research Repository at the Manitoba Center for Health Policy (MCHP). A subsample of 1,993 students was used in the SDQ validation study, and a full sample of 4,676 students was used for PAX program evaluation.

Statistical Analyses: The association between SDQ symptom scores and mental disorder was analyzed, and the screening properties of the SDQ were explored using multiple criteria including sensitivity, specificity, accuracy, and the receiver operating characteristic (ROC) value. We explored the best cut-off values of SDQ scores or classification for program evaluations by assessing all possible combinations of the SDQ scores. Then, a series of linear mixed models or generalized linear mixed models were fitted to illustrate how the SDQ scores or classification based on SDQ scores can be used for program evaluations.

Result and Conclusion: The SDQ symptom scores differ significantly among those with mental disorders and those without. Our proposed cut-offs for classifying as normal and abnormal showed better sensitivity than current ones. Using both the SDQ difficulty score and SDQ impact score had a better prediction for identifying mental disorders than using the SDQ difficulty score alone. Similarly, using both SDQ hyperactivity score and SDQ impact score have a better prediction for identifying ADHD than using the SDQ hyperactivity score alone. Using the SDQ symptom scores as continuous variables and dichotomized variables with SDQ classifications can lead to different estimations of program effect size and identify different factors associated with mental wellness.

Impact and Significance: The current study made inferences on the screening properties of SDQ for mental disorders of school-aged children in Canada, which fills a significant gap in current knowledge of the use of SDQ for mental health screening. The recommendation on the use of SDQ will allow us to identify children who may need early intervention and additional supports to promote healthy development. The evidence from this thesis helps mental health prevention program evaluation and understand what factors might facilitate the implementation of mental health programs in Canada.

Acknowledgment

Foremost, I cannot express enough thanks to my supervisor, Dr. Depeng Jiang for his invaluable guidance and tremendous support to get me this far. Aside from the training he provided in biostatistics, he strongly encourages me to enrich my experience from many angles and help me to reach my full potentials. I deeply appreciate his knowledge, kindness, flexibility, and patience.

I would like to thank my committee members. Special thanks to Dr. Joy Wei, who supervised me at Manitoba Health and guided me through not only epidemiological training but also helped me to adjust to the work environment and provides academic support. Many thanks to Dr. Shahin Shooshtari for being my committee member and sharing her insightful suggestions and providing prompt feedback.

I am extremely grateful for my selfless parents, loving sister and brother for their sacrifice and/or understanding that enables me to live all my first part of life in school worry-free and come this far. I am thankful to my loving husband, Yang Zhang; he has been particularly patient and amusing during the stressed time, made me being his proud and supported me with his gifted spicy dishes. I am thankful to my dear friends, Yixiu Liu and Dr. Shujun Huang for their understanding and company to help me through all the time I struggled and for being with me for every moment of my achievement.

I also appreciate much help from and the time spent with my “academic sisters”: Xuan Chen and Kun Liu. Thanks to Beibei Lyu, Xugang Zhang, Ou Lyu, and Colin Zhong for their social competence and their carnivore feast for warming up my stomach and heart.

Data used in this study are from the Manitoba Population Research Data Repository housed at the Manitoba Centre for Health Policy (MCHP) and were derived from data provided by

Manitoba Health, Seniors and Active Living, Manitoba Families, and Manitoba Education. I would like to acknowledges their permission for data usage. The results and conclusions are those of the author and no official endorsement by them is intended or should be inferred.

Last but not the least, thanks to the Research Manitoba Health Service Program for providing the financial support for me to work as a research assistant, and the George and Fayee Centre for Healthcare Innovation (CHI) for the opportunity of working as a junior consultant at CHI!

Chapter 1. Introduction

Approximately, 1.2 million Canadian children and youth (20%) are suffering from mental health problems and illnesses(Mental Health Commission of Canada, n.d.-a). Though the percentage of mental disorders among children and youth aged 5 to 24 remained unchanged over the period from 2007 to 2017, the hospital service use for mental disorders has increased substantially. In Appendix A-1, 61% and 60% of increases in visits to the emergency department and hospitalizations related to mental disorder were observed from 2008-2009 to 2018-2019, while obvious declines (26%) were observed in hospitalizations related to other conditions of this age group(Canadian Institute for Health Information, 2020). Moreover, the increase in medication-dispensation for mood and/or anxiety control and antipsychotic treatment is not small. According to the Canadian Institute for Health Information (CIHI, 2020)'s report, 9.1% of youth received mood/anxiety medication in 2018-2019. Even though there is a substantial amount of pharmaceutical use, no more than 20% of Canadian children and youth with mental illness actually receive appropriate treatment (Mental Health Commission of Canada, n.d.-a). In addition, Youth Mental Health Canada (YMHC) adds that the youth with mental conditions and crises need to wait more than a year on average to receive mental services and supports (n.d.). More threatening, mental illness is the top leading cause of disability and other harms(Youth Mental Health Canada, n.d.); in particular, about 24% of deaths among Canadian youth aged 15-24 are due to suicide each year(Youth Mental Health Canada, n.d.).

Most mental illnesses and disorders start in childhood and are extended to or become worse in adolescence and adulthood such as anxiety disorder, attention deficit hyperactivity disorder (ADHD), conduct disorder, and depression(Mental Health Commission of Canada, n.d.-b, pp. 7–8; Waddell, 2017, p. 25). The influences of mental illness and disorder are profound, and this adverse effect can be either short-term or long-term (Kutcher et al., 2010, p. 10; Waddell, 2017, p. 26). At the individual level, poor mental health increases the risks of physical illness and the likelihood of shortened life expectancy(Kutcher et al., 2010, p. 10). Mental illness and problems can also result in more than one possible social issues. Homelessness is an example of social concern faced by people having mental disorders in Canada(Smetanin et al., 2011, p. 51). As reported by Parliamentary Information and Research Service, over 66% of Toronto homeless people had a diagnosis for mental-illness at some point in their life(Smetanin et al., 2011, p. 51). Financially speaking, the Mental Health Commission of Canada reported that over 42.3-billion-dollar financial cost was directly attributable to public mental illness in 2011(n.d.-b). An even higher mental health-related healthcare cost, 290.9 billion dollars, is anticipated for 2041 if the cost curve is not to be bent actively (Mental Health Commission of Canada, n.d.-b). Obviously, there is an urgent need for strategic planning towards better population mental health, especially targeting Canadian children and youth.

Donovan and Spence (2000) showed that early-intervention and mental-disorder-prevention strategies are far more effective and cost-efficient than clinical treatments after diagnosable disorders have already been developed. However, attending medical appointments and get clinical assessments of mental health for children requires a large amount of time and effect for all involving parties (e.g. children, parents/guardians, and medical practitioners) and could be less accessible for children living in rural areas. Therefore, shorter screening tools are useful first steps in identifying children in need of further specialized assessments. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a widely-used assessment tool for mental health status. The SDQ and its scoring metrics provide the potentials to conduct early-stage mental health screening with easy access, simple administrative manners and low costs

(YouthinMind, n.d.-b). Several screening cut-offs using SDQ have been developed in Britain, Sweden, and the United States (Goodman et al., 2000; Arnfred et al., 2019; Shojaei et al., 2009). Though plenty of Canadian studies have leveraged this questionnaire, its predictive ability has not been studied in Canada (YouthInMind, n.d.). Hence, the overall purpose of this study is to develop the knowledge of using the SDQ for screening mental health problems among children and its implication on the evaluation of the mental health promotion programs.

Chapter 2. Literature Review

This chapter provides an overview of existing literature on SDQ as a screening tool of mental health and mental health prevention program. Firstly, we introduce SDQ and outline its development/limitations in scoring metrics and cut-offs. Following that, the administrative database and the Gold Standard are discussed to examine the outlined metrics and cut-offs. Besides, we report the previous findings and debates of the early mental health promotion programs for their significant role in our study. Finally, we discuss the knowledge gaps in the literature and objectives this thesis.

2.1 The Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) was first outlined for psychiatric screening by Robert Goodman in 1997 based on a British study (Goodman, 1997). This psychological assessment tool is widely recognized and frequently used in over 100 countries including Canada (YouthInMind, n.d.). More than 4,000 studies and 5,000,000 tests were performed with the help of the SDQ due to its structural briefness, easiness to use, affordability, flexibility in multiple versions, accessibility in different languages and strong study validity (YouthinMind, n.d.-b). In Canada, there are also 147 publications reported using SDQ according to the searchable database on the SDQ website (YouthInMind, n.d.).

The questionnaire is only two-page in length and consists of five major components. Specifically, four components evaluate the difficulties in emotion symptoms, conduct problems, hyperactivity-inattention, peer relationship problems, while one subscale quantifies the wellness in prosocial behavior (Goodman, 1997; see Appendix C). Each psychiatric-attribute-measuring component is an integration of five itemized questions with answers of “Not True”, “Somewhat True” and “Certainly True”. Overall, there are 25 items to measure different psychological features along with an impact supplement (Q26–30, see the second page of Appendix C), which provides extra insights into a child's overall distress and impairment measured by the responder of the questionnaire (Goodman, 1999).

Besides, the SDQ further supports the flexibility of having various roles of informants; in other words, any combination of children themselves, their parents and teachers can contribute their knowledge on the study participants depending on the study design and age specifics. Also, the structure of the SDQ is well studied and validated by many studies. For instance, Woerner (2004) performed validation using two approaches; one method compared the means of scales of community sample and clinically diagnosed sample and found that the total difficulty scores using the SDQ are significantly different for two samples. In another approach, Woerner treated the Development and Well-Being Assessment (DAWBA) [1] as a gold standard and found that the test results of using the SDQ and the DAWBA are consistent (2004). Therefore, the SDQ with effective mental-health-state categorizing methods provides the potential to conduct mental health screening.

2.2 The SDQ Scoring Metrics and Cut-off Level for SDQ Score

To use the questionnaire for screening, scoring metrics and cut-off levels are required. In terms of how to score the questionnaire based on filled-out forms, Table 2-1 displays the typical symptomatic metrics for children and youth in the age range 4 to 17 (Youthinmind, 2015). Notice that each subscale score is summed up with the scores of five associated items to a range from 0 to 10. As a result, the total difficulty score is simply the aggregation of emotion, conduct,

[1] Development and Well-Being Assessment (DAWBA) is another mental health assessment tool containing questionnaires, interviews and ratings. This tool covers a bigger age range (2-65-year olds) and identifies individuals with emotional, behavioral and hyperactivity disorders consistent with ICD-10 and DSM-IV or DSM-5 psychiatric diagnoses;
(YouthinMind, n.d.-a)

hyperactivity, and peer problem scores (ranged from 0 to 40), and the strength score is equal to the prosocial behavior scores (ranged from 0 to 10). In addition, Table 2-2 displays the scoring metrics for the impact supplement evaluation.

Table 2-1. Scoring Symptom Scores on the SDQ for 4–17 Year Olds

	Not True	Somewhat True	Certainly True
Emotional problems scale			
ITEM 3: Often complains of headaches… *(I get a lot of headaches…)*	0	1	2
ITEM 8: Many worries… *(I worry a lot)*	0	1	2
ITEM 13: Often unhappy, downhearted… *(I am often unhappy….)*	0	1	2
ITEM 16: Nervous or clingy in new situations… *(I am nervous in new situations…)*	0	1	2
ITEM 24: Many fears, easily scared *(I have many fears…)*	0	1	2
Conduct problems Scale			
ITEM 5: Often has temper tantrums or hot tempers *(I get very angry)*	0	1	2
ITEM 7: Generally obedient… *(I usually do as I am told)*	2	1	0
ITEM 12: Often fights with other children… *(I fight a lot)*	0	1	2
ITEM 18: Often lies or cheats *(I am often accused of lying or cheating)*	0	1	2
ITEM 22: Steals from home, school or elsewhere *(I take things that are not mine)*	0	1	2
Hyperactivity scale			
ITEM 2: Restless, overactive… *(I am restless…)*	0	1	2
ITEM 10: Constantly fidgeting or squirming *(I am constantly fidgeting….)*	0	1	2
ITEM 15: Easily distracted, concentration wanders *(I am easily distracted)*	0	1	2
ITEM 21: Thinks things out before acting *(I think before I do things)*	2	1	0
ITEM 25: Sees tasks through to the end… *(I finish the work I am doing)*	2	1	0
Peer problems scale			
ITEM 6: Rather solitary, tends to play alone *(I am usually on my own)*	0	1	2
ITEM 11: Has at least one good friend *(I have one goof friend or more)*	2	1	0
ITEM 14: Generally liked by other children *(Other people my age generally like me)*	2	1	0
ITEM 19: Picked on or bullied by other children… *(Other children or young people pick on me)*	0	1	2
ITEM 23: Gets on better with adults than with other children *(I get on better with adults than with people my age)*	0	1	2
Prosocial scale			
ITEM 1: Considerate of other people's feelings *(I try to be nice to other people)*	0	1	2
ITEM 4: Shares readily with other children… *(I usually share with others)*	0	1	2
ITEM 9: Helpful if someone is hurt… *(I am helpful is someone is hurt…)*	0	1	2
ITEM 17: Kind to younger children *(I am kind to younger children)*	0	1	2
ITEM 20: Often volunteers to help others… *(I often volunteer to help others)*	0	1	2

Source: YouthInMind

Table 2-2. Scoring the SDQ impact supplement

	Not at all	Only a little	A medium amount	A great deal
Teacher report:				
Difficulties upset or distress child	0	0	1	2
Interfere with PEER RELATIONS	0	0	1	2
Interfere with CLASSROOM LEARNING	0	0	1	2
Self-report report:				
Difficulties upset or distress child	0	0	1	2
Interfere with HOME LIFE	0	0	1	2
Interfere with FRIENDSHIPS	0	0	1	2
Interfere with CLASSROOM LEARNING	0	0	1	2
Interfere with LEISURE ACTIVITIES	0	0	1	2

Source: YouthInMind

In the following sections, we will discuss the detailed steps of classifying a child's mental health based on his/her calculated SDQ scores and examine the goodness of the classification.

2.2.1 Previous Studies on Cut-off Point

The pioneer researcher, Robert Goodman, did not only invent the SDQ but also outlined associated categorizing levels for the mental health scores (see Table 2-3; Goodman, 1997; Youthinmind, 2015). Three years later, he strived to improve the detection of child mental disorder within community settings. Meanwhile, the sensitivity of the SDQ diagnoses by a multi-informant approach was discovered to be better than the single-informant approach(Goodman et al., 2000). He reported that the specificity of multi-informant SDQs for 5-15 age group was 94.6%, while the multi-informant SDQ's sensitivity is no more than 64% (Goodman et al., 2000). Goodman, Renfrew, and Mullick (2000) further added that the chance-corrected agreement between the psychiatric diagnoses-rating predictions using the SDQ ("unlikely", "possible", or "probable") and clinical diagonal result exists. Likewise, a recent study in Denmark spotted gender differences in certain SDQ scores and tried to improve the cut-points by including gender-specific thresholds for categorization (see Table 2-4; Arnfred et al., 2019).

Table 2-3. Teacher-Completed-SDQ Categorization Thresholds Outlined by Goodman (Aged 4-17)

	Normal	Borderline	Abnormal
Total difficulty score	0-11	12-15	16-40
Emotional problem score	0-4	5	6-10
Conduct problem score	0-2	3	4-10
Hyperactivity score	0-5	6	7-10
Peer problems score	0-3	4	5-10
Prosocial score	6-10	5	0-4

Table 2-4. Gender-Specific and Teacher-Completed-SDQ Categorization Thresholds Outlined by Danish studies (Age 6-10 years)

		Close to average	Slightly raised/ Slightly lowered	High/Low	Very high/ Ver low
Boy	**Total difficulty score**	0-13	14-17	18-20	21-40
	Emotional problem score	0-3	4	5	6-10
	Conduct problem score	0-3	4	5-6	7-10
	Hyperactivity score	0-7	8-9	10	N.A.
	Peer problems score	0-2	3-4	5	6-10
	Prosocial score	5-10	4	3	0-2
Girl	**Total difficulty score**	0-8	9-13	14-16	17-40
	Emotional problem score	0-3	4-5	6	7-10
	Conduct problem score	0-1	2	3-4	5-10
	Hyperactivity score	0-3	4-5	6-7	8-10
	Peer problems score	0-2	3	4	5-10
	Prosocial score	7-10	6	5	0-4

2.2.2 Validation Measures and Finding Optimal Cut-off

Although the classification thresholds have been studied in the UK, Denmark and other countries (e.g. France and the United States; Shojaei et al., 2009; see Appendix B-1), the SDQ categorization and its predictive ability have not been examined in Canada. Consequently, there is a need to validate whether these existing cut-off levels are applicable to the Canadian children. Specifically, the following two questions are addressed in this thesis. Are there optimal/better cut-off values to use with SDQ in Canada? Does a combination of scores from the SDQ scales including the impact scale yield additional predictive value?

The accuracy of classification using cut-offs is usually evaluated using sensitivity, specificity, and overall accuracy (Pepe, 2003; Yuan et al., 2015). Sensitivity is defined as the proportion of correctly identified positive cases by a diagnostic test or model prediction, while specificity corresponds to the proportion of correctly identified negative cases (Silva et al., 2015). Furthermore, the overall accuracy quantifies the proportion of correctly classified cases out of all study subjects(Sammut & Webb, 2011). As discussed, different measures concentrate on different traits, and thereby inconsistent optimal categorizing thresholds can be achieved using the aforementioned statistical evaluation criteria. That is when the receiver operating characteristic (ROC) curve works toward our demand.

The ROC curve plots 1-specificity on the x-axis and sensitivity on the y-axis and thereby allows a comprehensive tradeoff between sensitivity and specificity(Yuan et al., 2015). This curve shows possible points of cut-off levels and lets researchers decide which one or more points optimize the diagnostic accuracy. There are several developed and commonly-used decision criteria that help to find better cut-points using ROC analysis. In the first approach, the optimal cut-off is determined by maximizing the Youden index (J), which quantifies the difference between sensitivity and false-positive rate over all possible cut-off thresholds (Fluss et al., 2005; Perkins & Schisterman, 2005; Unal, 2017). Alternatively, Unal (2017), Perkins and Schisterman (2006) proposed that the optimal threshold can also be achieved by finding the closest point of diagnostic tests from the top-left corner of the ROC plane. Conclusively, the ROC curve can be used to not only validate particular test results but also to propose a better cut-off level. Still, decisions of emphasizing one or more diagnostic criteria are open to discussion depending on the study's needs of higher sensitivity or specificity.

Besides, if we think from the healthcare cost perspective, one of our goals is to bend the cost curve for the mental conditions developed in early childhood. As far as Kutcher et al. (2010) and Donovan and Spence (2000) as concerned, early treatments or prevention are more cost-effective than delayed treatment after mental disorders have been developed and diagnosed in adulthood ($c_{t_1} < c_{t_2}$, where c_{t_1} indicates the cost associated with early treatment, and c_{t_2} indicates the cost associated with delayed treatment costs). Moreover, we would also take the additional assessment cost (c_a) resulted from false positive detection based on our cutoff into consideration. With developing a comprehensively and functional classification tool, the aggregated healthcare cost accounting for all three components should be minimized.

2.3 Administrative Database and the Gold Standard

The administrative health database is a place to hold the regularly-collected administrative data on health care programs like medical care insurance, visits to clinics, hospitals, laboratories or pharmacies (Manitoba Centre for Health Policy, n.d.-b). This type of dataset is usually considered as an inexpensive tool for identifying the diseases or conditions for a large and representative group of patients (Hashimoto et al., 2014; Sarrazin & Rosenthal, 2012). Using the administrative health data, the researchers can retrieve information on medical diagnosis, accepted coding system, hospital administration and stay, and severe outcome related to the disease of interest (Hashimoto et al., 2014; Romano et al., 2010).

Here in Manitoba, there is such a database hosted in the Manitoba Centre for Health Policy (MCHP). This data centre receives, holds, and provides access to cleaned multi-dimensional administrative health and non-health data (Manitoba Centre for Health Policy, n.d.-a). It provides the opportunity to validate the screening tool (the SDQ) and examine its predictive ability of SDQ.

Speaking of how to validate an existing or proposed cut-off level, one cannot avoid mentioning a gold standard to compare with. A gold standard in the diagnostic testing procedure often means a typical time-honored or well-known method of identifying the diseases or conditions accurately(Claassen, 2005). Nonetheless, the gold standard test/procedure can be usually associated with unignorable cost, cumbersome processes or invasive risks(Glen, 2017). In our context, the gold standard for comparison is defined as whether the children have particular mental conditions based on the case definition summarized by the Manitoba Centre for Health Policy (details will be provided in Section 0 and Appendix B-2). Obviously, getting diagnosed with mental conditions by physicians or mental health specialist is more expensive, and takes more effort for the child's family and healthcare system than filling out a survey. Thus, it is quite meaningful and rewarding to study the validity of the SDQ.

However, the mental outcomes identified using the administrative database have limited capacity to classify individuals into specific groups having mental difficulties (i.e. borderline v.s. abnormal / slightly raised, high, v.s. very high). In addition, no former literature has examined how to classify individuals with different levels of mental outcomes based on the available disorders identified from the database. Therefore, in this study, we concentrated on the discriminative capacity of SDQ over children with or without mental disorders and concatenated the cut-offs of previous studies into two classification bands.

Table 2-5. Concatenated Teacher-Completed-SDQ Categorization Thresholds Outlined by Goodman (Aged 4-17)

	Normal	**Abnormal**
Total difficulty score	0-11	12-40
Emotional problem score	0-4	5-10
Conduct problem score	0-2	3-10
Hyperactivity score	0-5	6-10
Peer problems score	0-3	4-10

Prosocial score	6-10	0-5

Table 2-6. Concatenated Gender-Specific and Teacher-Completed-SDQ Categorization Thresholds Outlined by Danish studies (Age 6-10 years)

		Normal	Abnormal
Boy	**Total difficulty score**	0-13	14-40
	Emotional problem score	0-3	4-10
	Conduct problem score	0-3	4-10
	Hyperactivity score	0-7	8-10
	Peer problems score	0-2	3-10
	Prosocial score	5-10	0-4
Girl	**Total difficulty score**	0-8	9-40
	Emotional problem score	0-3	4-10
	Conduct problem score	0-1	2-10
	Hyperactivity score	0-3	4-10
	Peer problems score	0-2	3-10
	Prosocial score	7-10	0-6

2.4 Early Mental Health Promotion Program

Mental health promotion describes efforts made for advancing population mental health and wellbeing (Commonwealth Department of Health and Aged Care, 2000, Chapter Introduction), and mental health prevention describes the proactive actions taken before the onset of mental illness(World Health Organization, 2002). Mental health promotion and prevention programs incorporate the respective positive life strategies and skills to achieve the promotive and/or preventative mental health goals. The idea of prevention programs was first pointed out over a century ago, but the researchers did not apply it in real studies until the 1950s (Ojemann et al., 1955). In 1997, Durlak and Wells conducted a meta-analysis on 177 primary mental health programs for preventing children from developing social and behavioral difficulties. They found that most intervention programs showed statistical significance in improved competencies and decreased mental problems in participants(Durlak & Wells, 1997). Later on, Neil and Christensen specifically reviewed the effectiveness of 27 school-based prevention and early intervention programs, and they commented that 21 out of all examined programs showed positive results in reducing symptoms of anxiety with statistical significance (effect size: 0.11–1.37), while the rest of the programs did not reveal significant results(2009). Moreover, Weare and Nind (2011) reviewed 52 systematic reviews and meta-analyses about this topic, and they confirmed that the early interventions overall benefit the children and youth mentally, socially and academically. However, a number of studies failed to include a follow-up measurement to examine the long-term effect and may result in biases in some program effect estimations according to Neil and Christensen's arguments (2009).

PAX, also called the PAX Good Behavior Game (PAX GBG), is one example of such school-based early mental health promotion programs. This program provides a set of training to teachers and students in active teaching and self-regulation skills based on ample evidence, including research results from the Good Behavior Game (GBG) and the PeaceBuilders program(Government of Manitoba, Healthy Child Manitoba Office, n.d.-a). For instance, the GBG has shown reductions in emotional/behavioral problems, externalizing behavior symptoms, ADHD symptoms, suicidal ideation/attempts, and substance use, as well as advancement in academic performance and high school graduation(Bradshaw et al., 2009; Brownell et al., 2018; Ialongo et al., 2001; Kellam et al., 2008; Poduska et al., 2008; Washington State Institute for Public Policy, 2018, sec. Meta-analysis). PAX, as a more advanced and integrated model of two successful programs, is supposed to increase the likelihood of students to succeed academically, socially and economically in the long run.

2.5 Purpose and Research Objectives

In summary, the purposes of this thesis are to investigate the screening properties of the SDQ for the mental disorder of Canadian children and examine the implication of mental health promotion program evaluation of using this screening tool. Explicitly, the research objectives are: (1) to study the discriminative capacity of the SDQ scales of children with and without mental disorder; (2) to investigate the applicability of existing SDQ cut-off values developed in the UK and Demark to Canada, and explore better or optimal SDQ cut-off thresholds, which can be recommended for Canada; (3) to investigate whether combining the SDQ scale values and Impact scale advances the screening capacity; and (4) to examine the effectiveness of the PAX program using two different approaches: one with SDQ scale scores treated as continuous outcomes and one with the SDQ scale categories (with cut-off levels) used as categorical outcomes.

Chapter 3. Study Design and Samples

This thesis uses the linked administrative health databases and PAX program data. In particular, we use three data sources: the PAX Good Behavior Game RCT pre- and post-mental health outcome measures using SDQ collected in 2011/2012, the Manitoba Grade 5 Mental Health Survey (G5 Survey) in 2015/2016, and health, social services administrative data up to 2017 held in the Population Health Research Repository at the MCHP. In this chapter, we present information on data sources and outcome measures used in this thesis. We first describe the designs of the PAX RCT pilot study and the G5 Survey. Then, introductions are given regarding the samples retrieved from both studies and the measures used to quantify our research interests. At the end of this chapter, we describe the ethical considerations taken to ensure the privacy of the personal information and the proper data handling.

3.1 Study Designs

3.1.1 PAX Good Behavior Game (PAX GBG) in Manitoba

In the 2011/12 fiscal year, the Healthy Child Manitoba Office (HCMO) initiated the first province-wide cluster randomized controlled trial[2] to invite all schools to participate in the pilot PAX program (Government of Manitoba, Healthy Child Manitoba Office, n.d.-a). The invitation was delivered to a large diversity of school types (including public, private, First Nation and independent schools) with an intent to cover all school divisions (Government of Manitoba, Healthy Child Manitoba Office, n.d.-a). In the end, 196 schools from all regions of the province accepted the invitation to the PAX pilot study including First Nation and independent schools. Randomization is performed at the school level; in other words, teachers of 50% of the schools

---

[2] Cluster randomized controlled trial (RCT) is the randomized controlled trial where treatments are allocated to groups of individuals (Bland, 2004).

were provided with two days of PAX training and implemented the program in their class in 2011/12 school year, and the other 50% received training and implemented PAX in the following school year (waitlist control group; Brownell et al., 2018; Government of Manitoba & Healthy Child Manitoba Office, n.d.-a).

Before and after the PAX implementation, students' mental wellness data were collected using SDQ and linked with health, education and social services administrative data held in the Population Health Research Repository at the MCHP. As part of a project funded by Manitoba Health, Seniors and Active Living (MHSAL), MCHP published a deliverable report in 2018. Based on this report, there are 4,676 valid records for the grade 1 students involved in the PAX program with 2,576 students assigned to the PAX/exposed group and 2100 students in the control group after removing duplicates, students not enrolled in grade 1 and students with invalid linking identifiers. Notice that the identifier at the student level refers to a scrambled version of the Personal Health Identification Number (PHIN), which is de-identified but provides possible linkage across different datasets (Brownell et al., 2018).

3.1.2 The Grade 5 Mental Health Survey (G5 Survey)

HCMO launched the Grade 5 Mental Health Survey in May 2016 (n.d.-c). The G5 Survey is Manitoba's first-ever province-wide survey to measure the mental health of over 12,000 Grade 5 students (Government of Manitoba, Healthy Child Manitoba Office, n.d.-c). HCMO implemented the G5 Survey to measure youth's mental health status and provide support to youth and children in Manitoba accordingly (Government of Manitoba, Healthy Child Manitoba Office, n.d.-b). The G5 survey was also conducted using SDQ, which gives us the opportunity to follow up with the majority of students in the PAX program. In specific, we found 3,302 out of

4,676 students in the PAX pilot study were also involved in the G5 Survey after linking data from both programs.

3.2 Study Sample

Our primary goal is to investigate the screening properties of the SDQ for the mental disorder of Canadian children. For this (Objective 1 to 3), our analyses include 1,993 students who have complete data in all SDQ components and the gender at the post-PAX time point. Amongst them, about 13% of students had mental disorders, and the proportion of mental disorders among the whole cohort (n=4,676) is also 13%. Our secondary goal (Objective 4) is to examine the implication of mental health program evaluation using SDQ as a screening tool, and we are able to include all 4,676 students' data who participated in the PAX pilot RCT study. For this objective, we included all individuals who participated in the PAX pilot study even they might have missing data at follow-up assessments. More sample details are provided in the result section.

3.3 Study Measures

3.3.1 Measures of Mental Health

3.3.1.1 Mental Health Definition

Mental health is a state of wellbeing that is a measure of self-awareness, the ability to handle life stresses, work productivity, and social activities. Good mental health gives protection to people from life stress and possibly reduces the chances of developing mental illness and/or disorders (Zaun et al., 2016). Mental disorder is psychiatric health conditions related to how people think, perceive and behave (American Psychiatric Association, 2018). Such disorder/illness is common and can impair one's personal functioning in social, work and/or family life (American

Psychiatric Association, 2018). To assess one's mental status, mental health indicators are used (Mental Health Commission of Canada, 2015, p. 5) and introduced in the following sections.

3.3.1.2 Mental Health Measures using the SDQ

Teachers involved in the PAX program and the G5 Survey filled out the SDQ for their students, and the students also filled out the G5 Survey themselves. Specifically for the PAX pilot study, the teachers were required to fill out the questionnaires for students at two time-points: pre-PAX (at the beginning of 2012) and post-PAX (June 30, 2012). Teachers and students involved in the G5 Survey completed the SDQ during 2016 (Government of Manitoba, Healthy Child Manitoba Office, n.d.-c). The study variables retrieved using SDQ were scores for various questionnaire components, including scores for each subscale (ranged 0–10), total difficulty score (ranged 0–40) and the impact scale score (ranged 0–3).

3.3.1.3 Mental Outcome/Indicator Definitions

Mental outcomes for our study cohort were evaluated at four different angles by five measurements. These mental disorders we considered are attention deficit hyperactivity disorder (ADHD), conduct disorder, mood and anxiety disorder and emotional and behavioral disorder. These outcomes were extracted from several datasets hosted at the Manitoba Centre for Health Policy (MCHP), including the Hospital Abstracts dataset, the Medical Claims/Services dataset, the Drug Program Information Network dataset, the Manitoba Health Insurance Registry dataset, the Social Assistance Management Information Network dataset, and the Child and Family Services dataset(Brownell et al., 2018). According to the mental disorder definition outlined in the MCHP report (see Appendix B-2), we can link our study participants with the administrative data repository and detect the presence of interested mental outcomes including diagnoses of

ADHD, conduct disorder or mood and anxiety disorder, prescriptions for ADHD or funding assistance for emotional and behavioral disorders (Chartier et al., 2016, Chapter 2). All these outcome indicators fall in binary categories of either “1” representing there is a history of diagnosis or “0” representing there is no such history.

Alike the SDQ measurement of mental health was performed at different points of time, the mental outcomes of the PAX children were extracted for comparable time frames. The pre-PAX time frame is defined as the service-recipient year between January 1, 2000 and December 31, 2011 for those aged between three and five years old; during-PAX mental outcomes are extracted between January 1, 2012 and June 30, 2012 for those aged over three years old; and post-PAX outcomes are extracted between July 1, 2012 and March 31, 2017. After retrieving each disorder-identifying outcome of an individual child for the desired time frame, we sum up different outcomes to one aggregated variable named mental disorder or mental outcome, which ranged from 0 to 5. Then, we dichotomize children corresponding to the mental health classification bands discussed for SDQ: “Normal” for children having zero outcomes of all five disorders and “Abnormal” for children having one or more outcomes of interests.

3.3.2 Other Study Measures

Besides the measures of mental health, there are several other variables considered by our research goals. For the prevention program evaluation, we include gender, residency status (urban or rural residence), and socioeconomic status in our analyses. The residency status and socioeconomic status were extracted from the Canadian Census Data. The residency status distinguishes the residence in urban (living in Winnipeg or Brandon) or rural areas (living elsewhere in Manitoba; Manitoba Centre for Health Policy, 2002, 2020), while the socioeconomic status reflects non-medical, regional-level, social determinants of health, which is quantified by the Socioeconomic Factor Index (SEFI) developed by MCHP (Metge et al., 2009). Because the receipt of family-supporting services may have an impact on student's mental wellness, the following variables are deemed valuable to our study and extracted using the Manitoba Families data. They are the status of whether the student is a Child in Care (CIC) in Child and Family Services, the receipt of Voluntary Family Services (VFS), and the receipt of Income Assistance (IA) by the student's family.

3.4 Ethical Considerations

All the data used in the analysis was administrative data. As the standard protocol, the data were anonymized by HCMO. The ethical approvals from the University of Manitoba's Human Research Ethics Board (HS21450 (H2018:016) and the Manitoba Health's Health Information Privacy Committee (HIPC No. 2017/2018 - 69) were obtained, as well as the data sharing agreement with HCMO for the use of data housed in HCMO. Regarding the confidentiality and security, all the analyses were done in HCMO and no linkage was required for the current analyses.

This study uses the administrative data housed at MCHP. As per MCHP regulation, the data were de-identified and linked together using the unique scrambled Personal Health Identification Numbers (SCRPHIN). To access the data, we acquired approval from the University of Manitoba Health Research Ethics Board (HREB: HS23564 (H2020:013)); besides, this thesis is also a part of the approved research project led by Dr. Jiang, titled "The PAX Program in Manitoba: A Positive Approach to Promoting Mental Health and Wellbeing" (HREB: H2015:121; HIPC#: 2017/2018-09).To ensure the security of personal information, data were only accessed on the MCHP remote server, where login requires passing a combination of credential checks: password and security token pin that changes every minute. All data management, programming and analyses were performed using SAS® version 9.4 and R version 3.6.1 on the MCHP server. When reporting results, no names, addresses or summaries with frequency counts below five is disclosed to protect the privacy of personal information and the confidentiality of data.

Chapter 4. Statistical Analyses

In this chapter, we present statistical methods for each objective described in Chapter 2. Section 4.1 describes the methods to study the discriminative capacity of the SDQ scales. Section 4.2 describes the criteria and methods for the development of SDQ cut-offs. Sections 4.3 introduces the statistical methods, multi-level logistic regression model, to investigate how the SDQ impact scale can advance the screening capacity of SDQ scales. In Section 4.4, we describe the multilevel regression models for the PAX program evaluations.

4.1 Objective 1 – Discriminative Capacity of SDQ Scales

Preliminarily, violin plots display the distribution of selected variables. For our interests, visualizing the SDQ total difficulty score by the presence of mental disorders allows exploring whether there exists a tendency that students with mental disorders have higher difficulty scores. To draw a statistical conclusion on the distributional difference, t-tests were conducted to compare the group means for different SDQ components (i.e. subscale scores and total difficulty scores). In the next sub-section, analyses were extended to understand the place of cut-off thresholds on the distribution of the SDQ scores.

4.2 Objective 2 – SDQ Cutoffs for Screening Purpose

4.2.1 Validation Measurement

There are multiple measures available to quantify the appropriateness of classification thresholds, such as sensitivity (Se), specificity (Sp), and overall accuracy (ACC). As mentioned in Section 0, each measure represents a proportion of correct classification with a different numerator-denominator combination considered respectively. The formulae of the aforementioned statistics are listed in (1)-(3).

$$Se = \frac{Number\ of\ individuals\ with\ the\ disorder\ are\ correctly\ classified}{Number\ of\ individuals\ who\ have\ the\ disorder} \quad (1)$$

$$Sp = \frac{Number\ of\ individuals\ without\ the\ disorder\ are\ correctly\ classified}{Number\ of\ individuals\ who\ don't\ have\ the\ disorder} \quad (2)$$

$$ACC = \frac{Number\ of\ individuals\ who\ are\ correctly\ classified}{Total\ number\ of\ individuals\ in\ study} \quad (3)$$

We emphasize sensitivity although all three measures play important roles in our analyses since the essential purpose of this study is to screen out the children with mental conditions (i.e. the children whose mental health is in an abnormal state). To answer the question of which cut-off is optimal, either one of the three following approaches is selected to ensure not only achieving a high sensitivity value for detection but also generally maintaining a reasonably high specificity. They are (1) maximizing Youden's index or maximizing the sum of sensitivity and specificity, (2) minimizing the distance from the cut-off to the top-left corner of the ROC plane, and (3) minimizing the healthcare cost for decisions of classification.

Youden's index approach is trying to maximize the function, $sensitivity + specificity - 1$, which is essentially the same as maximizing the sum of sensitivity and specificity. Obviously, from the composition of the function, Youden's index accounts for the measures from two dimensions and ensures a fair amount of balance for classifying students.

The second approach emphasizes the distance of the discussed cut-off's position (1-specificity, sensitivity) to (0,1) on the Cartesian coordinate system, which is, in essence, measuring the tradeoff between sensitivity and specificity with quite similar logic of Youden's index method.

In the third method, healthcare costs related to each decision can also be treated as a penalty for making a false identification. By incorporating the costs in the optimization function, these monetary weights act as natural decision-making mechanism to achieve a threshold with the lowest financial burden caused by the early childhood mental complication. In Table 4-1, the average healthcare cost for each classified situation is assumed to be C_{t_1}, C_a, C_{t_2}, and 0. Hypothetical numerical cost will be used in the validation steps as no average mental health assessment fee and average mental condition treatment fee are available in the literature to the best of my knowledge.

Table 4-1. Costs Associated with Different Classification Decisions

		Mental Disorder Found in Administrative Database	
		Abnormal $x_i = 1$	**Normal** $x_i = 0$
Prediction by SDQ Total Difficulty Score	**Abnormal** $t_i = 1$	True Positive (TP) $Cost_{TP_i} = C_{t_1}$	False Positive (FP) $Cost_{FP_i} = C_a$
	Normal $t_i = 0$	False Negative (FN) $Cost_{FN_i} = C_{t_2}$	True Negative (TN) $Cost_{TN_i} = 0$

Notations:
- $\boldsymbol{C_{t_1}}$ denotes the treatment fee at detection, where the treatment fee is associated with the mental conditions detected at an early stage;
- $\boldsymbol{C_{t_2}}$ denotes the delayed treatment fee, where the mental condition is not found during the SDQ screening and results in delayed treatments after a student has grown up; and
- $\boldsymbol{C_a}$ denotes the cost associated with unnecessary follow-up, clinical assessment for mental disorder when a student actually does not have a mental condition.

Then, the total cost for a selected cut-off can be computed as follows:

$$Total\ Cost = \sum_{i=1}^{N} [x_i * t_i * C_{t_1} + x_i * (1 - t_i) * C_{t_2} + (1 - x_i) * t_i * C_a]$$

In an ideal case, where there is no false positivity nor false negativity, the total cost for this cohort of students should be reduced to

$$\begin{aligned} Total\ Cost &= \sum_{i=1}^{N} \left[x_i * t_i * C_{t_1} + x_i * (1 - 1) * C_{t_2} + (1 - 1) * t_i * C_a\right] \\ &= \sum_{i=1}^{N} x_i * t_i * C_{t_1} \end{aligned}$$

Noticed that if the delayed treatment cost is enormously higher than the early treatment cost, then the cut-off will be approaching a lower value (e.g. 1), which tries to flag nearly every student as abnormal and refers all students to clinical assessments for healthcare-cost-saving purposes.

4.2.2 Exhaustively Listing Possible Categorizing Thresholds

In this thesis, we used the exhaustive list approach to find better or optimal cut-offs. For instance, the SDQ total difficulty score is valued as integer numbers and ranges from 0 to 40. As a result, we can list out all possible thresholds to classify the outcomes into two levels (see example in Table 4-2). For the first trial, all students with SDQ total difficulty score 0 will be classified as “Normal”, all students scored 1+ will be classified as “Abnormal”. The same process will be carried on until we exhaust all combinations. Meanwhile, the performance of each combination of thresholds will be evaluated by comparing the aforementioned validation measures. Similar procedures are carried out for subscale scores. In the end, the analytical results imply us with one or more improved cut-off thresholds.

Table 4-2. Demonstration Example of Exhaustive List of Cut-off Combinations Based on Total Difficulty Scores

	Normal	**Abnormal**
Trial 1	0	1-40
Trial 2	0-1	2-40
Trial 3	0-2	3-40
…	…	…
Trial 39	0-39	40

4.3 Objective 3 – Use of the SDQ Score and Impact Scale

To examine the goodness of the cut-off developed using the exhaustive listing approach and further enhance the mental disorder prediction, the logistic regression model is considered as it allows the modeling of binary outcome and the inclusion of multiple covariates. In addition, the data are collected within different classes, which violates the independence assumption of observations. Thus, multi-level models or mixed effect models are used to compensate for the class-level variations. In this case, a two-level mixed model is a good fit for the underlying data structure for its ability to handle both the within-class and the between-class variations.

The combined logistic regression and two-level mixed model is represented in the following matrix formulation, and this model evaluates the probabilistic variation in mental disorder status by SDQ scores/categories.

To examine whether the use of SDQ scales or classification can help to predict mental health disorders and whether combining it with SDQ impact scale can improve the prediction, we fit hierarchical models with mental health disorders as the outcome. Since the outcome variable, mental disorder, is a binary outcome, logistic regression is used. As our data are from a cluster randomized controlled trial (RCT) with nested data structure (students nested within classrooms), multilevel logistic regression is used.

The multilevel logistic regression (also called random effect logistic regression or generalized linear mixed model) is represented in the following matrix formulation, and this model evaluates the probabilistic variation in mental disorder status by SDQ scores/categories.

$$log\left[\frac{P(Y'_{ij}=1)}{1-P(Y'_{ij}=1)}\right] = \boldsymbol{X_j \Gamma} + \boldsymbol{Z_j R_j}$$

Here, $\mathbf{X_j}$, and $\mathbf{Z_j}$ stand for the design matrices while $\mathbf{\Gamma}$, and $\mathbf{R_j}$ represent the fixed effects, the between-class random effects with respect to the design matrices. The outcome is a binary variable of mental disorder presences and denoted by $Y'_{\boldsymbol{ij}}$ for the student i in the classroom j.

4.4 Objective 4 – PAX Program Evaluation

To investigate the PAX program on reducing the difficulties, a series of three-level multilevel models (mixed effect models) is fitted with SDQ symptom scores assessed at multiple times or classification based on SDQ symptom scores as outcome variables. Section 4.4.1 presents the three-level mixed model for treating SDQ symptom scores as continuous outcomes while Section 4.4.2 presents the generalized three-level mixed model for the classifications based on the SDQ scores as categorical variables.

4.4.1 Mixed Effect Regression Model

A mixed model or random effect regression model takes both intra-personal effects over time and inter-personal effects into consideration. Besides, a three-level mixed effect model accounts for a higher-hierarchical cluster variation like students nested within classes. The composite matrix formulation of such a model is given as follows.

$$\boldsymbol{y_{ij}} = \boldsymbol{X_{ij}}\boldsymbol{\Gamma} + \boldsymbol{Z_{ij}}\boldsymbol{R_{ij}} + \boldsymbol{W_{ij}}\boldsymbol{U_j} + \boldsymbol{\varepsilon_{ij}} \tag{4}$$

Here, $\boldsymbol{X_{ij}}$, $\boldsymbol{Z_{ij}}$ and $\boldsymbol{W_{ij}}$ stand for the design matrices while $\boldsymbol{\Gamma}$, $\boldsymbol{R_{ij}}$ and $\boldsymbol{U_j}$ represent the fixed effects, the within-class random effects and between-class random effects with respect to the design matrices. The outcome matrix is denoted by $\boldsymbol{y_{ij}}$ for the student i in the classroom j. Note that the intra-personal errors ($\boldsymbol{\varepsilon_{ij}}$) are assumed to be uncorrelated with $(\boldsymbol{R_{ij}}, \boldsymbol{U_j})$ and follows an independent normal distribution with common variance, σ_{ε}^2; while $(\boldsymbol{R_{ij}}, \boldsymbol{U_j})$ follows multivariate normal distributions, and $\boldsymbol{R_{ij}}$ and $\boldsymbol{U_j}$ are assumed to be independent as well. Typically, a mixed effect model can also be composed in a multi-level form as described in equation **Error! Reference source not found.**)-**Error! Reference source not found.**.

Level 1: Intra-personal change over time (from pre-PAX to G5)

$$y_{tij} = \pi_{0ij} + \pi_{1ij} \cdot time + \varepsilon_{tij} \quad (5)$$

y_{tij} is the SDQ score (e.g. total difficulties) of the student i in the classroom j measured at time t; π_{0ij} is the average SDQ score of the student i in the classroom j at time 0 (i.e. before PAX implementation); π_{1ij} is the average change of the SDQ score per assessment for student i in the classroom j; and ε_{tij} is the residual.

Level 2: Inter-personal variability within a class (random intercept and random slope)

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + r_{0ij} \\ \pi_{1ij} &= \beta_{10j} + r_{1ij} \end{aligned} \quad (6)$$

β_{00j} is the initial j^{th}-class-average SDQ scores; β_{10j} is the average change over time of the class j; and r_{0ij} *and* r_{1ij} are the corresponding random individual effects.

Level 3: Variability among classes (random intercept and random slope)

$$\begin{aligned} \beta_{00j} &= \gamma_{000} + \gamma_{001} \cdot \text{PAX} + u_{00j} \\ \beta_{10j} &= \gamma_{100} + \gamma_{101} \cdot \text{PAX} + u_{10j} \end{aligned} \quad (7)$$

γ_{000} is the grand mean of all participants involved in the PAX program at pre-test; $\gamma_{001}$is the mean difference between the PAX group and wait-list control group; γ_{100} is the average change over the three time-points; γ_{101} is the difference of average change between the PAX group and the control group; and u_{00j} as well as u_{10j} are associated with random class effects.

4.4.2 Mixed Effect Logistic Regression

Sometimes, practitioners might be interested in how the intervention program affecting the likelihood of being abnormal or whether the proportion of students with abnormal mental health being reduced by the intervention. In that case, we would need to analyze the SDQ difficulty scores by dichotomizing them into binary variables (“Normal”, “Abnormal”). To study the program effects in disease prevalence reduction as discussed, three-level logistic regression models should be considered, as they can account for both the dichotomized outcome variable and the random effects among different students and classes. Moreover, using these hierarchical models advances our understanding of our proposed cut-off versus the previously outlined cut-offs, and factors that will increase/decrease the likelihood of being mentally abnormal. Then the adjustment or supports could be provided in accordance.

For the matrix representation, the models are extended above the two-level model discussed in the previous sub-section.

$$log\left[\frac{P\left(Y'_{tij}=1\right)}{1-P\left(Y'_{tij}=1\right)}\right]=\boldsymbol{X_{ij}}\boldsymbol{\Gamma}+\boldsymbol{Z_{ij}}\boldsymbol{R_{ij}}+\boldsymbol{W_{ij}}\boldsymbol{U_{j}} \quad (8)$$

Y'_{tij} is the classified mental health status of the student i in the classroom j at time t using optimal cut-off proposed.

Chapter 5. Results

In Chapter 5, we report results from the statistical analyses described in Chapter 4. The comparison of the SDQ-symptom-score distribution between those with mental disorders and those without is reported in Section 5.1. The screening properties of SDQ and our proposed new cutoffs of SDQ are reported in Section 5.2. In Section 5.3, we show results about how the SDQ symptom scores can predict mental disorders and whether using the SDQ impact scale can improve the prediction. The effects of the PAX program using SDQ symptom scores as continuous variables or classification based on symptom scores as dichotomized variables are reported in Section 5.4. In this section, we also report results about who are more likely to have abnormal mental health. A short summary of results is provided at the end of this chapter.

5.1 Objective 1 – Discriminative Capacity of SDQ Scales

The discriminative capacity of SDQ was explored by comparing the SDQ symptom scores between those with the mental disorders identified from the health administrative database and those without. We used the post-PAX SDQ symptom scores as an illustration example, and the complete sample at post-PAX is of size 1,993. Figure 5-1 overviews the distribution of the SDQ total difficulty scores between the students with and without mental disorders during the period of the post-PAX program. It is clear to see that the density of the normal student group clusters around the lower end of the scores, while the distribution for the student with mental disorder(s) is bi-modal and centers around 10–15. Preliminarily, a t-test was carried out to examine the difference between those with mental disorders and those without, and the results are shown in Table 5-1. It indicates that the average SDQ total difficulty score and subdomain scores are significantly higher among those with mental disorders than those without.

Figure 5-1. Violin Plot of the SDQ Total Difficulty Score by Mental Disorder

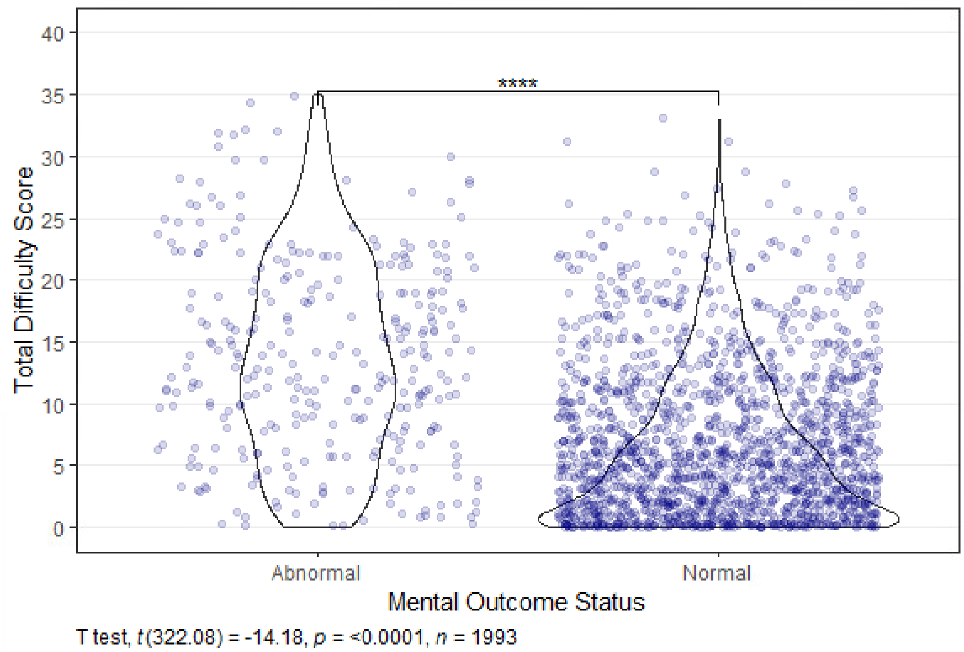


Table 5-1. Comparison of SDQ Score and Subscale between Different Outcome Groups

Variable	Normal (N=1,723) Mean (SD)	Abnormal (N=270) Mean (SD)	P-value	
Total difficulty score	6.0772 (6.1333)	13.1926 (7.8778)	<0.0001	***
Conduct problem score	0.9321 (1.6946)	2.6593 (2.7107)	<0.0001	***
Emotional problem score	1.2565 (1.8458)	2.1630 (2.4952)	<0.0001	***
Hyperactivity score	2.7667 (3.0155)	5.9407 (3.4226)	<0.0001	***
Peer problems score	1.1219 (1.7054)	2.4296 (2.2651)	<0.0001	***

Note: p-value <0.1 +, <0.05 *, p-value <0.01 **, p-value<0.001 ***

Furthermore, we examined whether the discriminative capacity of the SDQ symptom scores is different between males and females. Figure 5-2 shows the violin-plotted distributional comparisons with a gender-specific perspective, and Table 5-2 displays the summary of means and standard deviations of SDQ components by gender and mental disorders. It is noticed that male students tend to have higher SDQ difficulty scores than female students for those without mental disorders. Plus, there are gender differences in the SDQ difficulty score among those with mental disorders. Figure 5-2 also indicates that for both males and females, those with mental disorders have higher SDQ difficulty score than those without mental disorders.

Figure 5-2. Violin Plot of the SDQ Total Difficulty Score by Mental Disorder and Gender

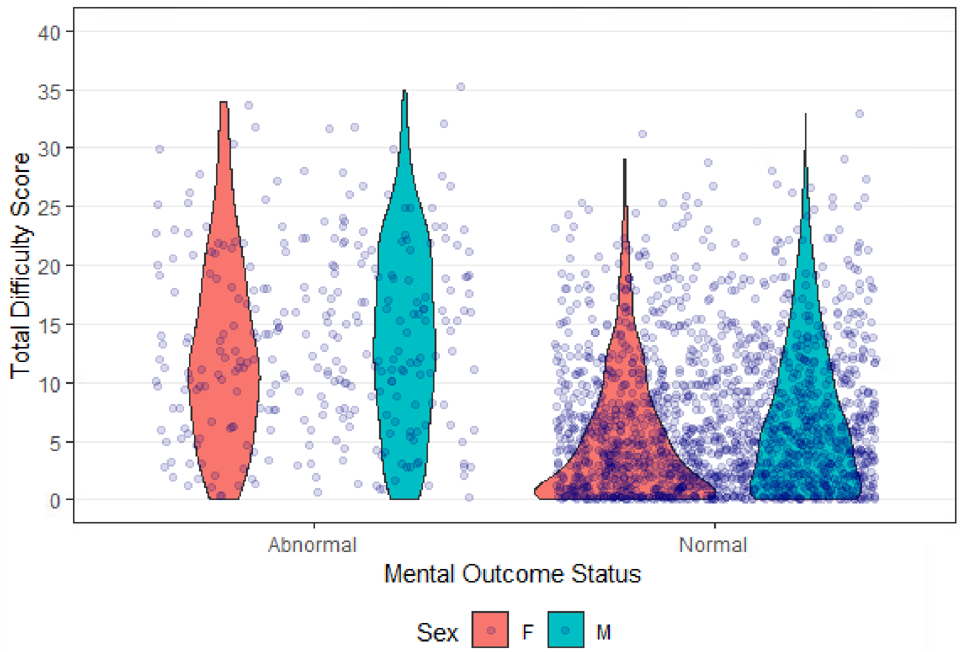


Table 5-2. Comparison of SDQ Score and Subscale between Different Outcome Groups

	Boy (N=986)		Girl (N=1,007)	
	Normal Mean (SD)	**Abnormal Mean (SD)**	**Normal Mean (SD)**	**Abnormal Mean (SD)**
Total difficulty score	7.1731 (6.5125)	13.6271 (7.9891)	5.1072 (5.6043)	12.3656 (7.6354)
Conduct problem score	1.1582 (1.8872)	2.8362 (2.6525)	0.7319 (1.4761)	2.3226 (2.8020)
Emotional problem score	1.2040 (1.7728)	1.9492 (2.4594)	1.3031 (1.9078)	2.5699 (2.5255)
Hyperactivity score	3.5612 (3.2364)	6.4011 (3.3528)	2.0635 (2.6127)	5.0645 (3.4001)
Peer problems score	1.2497 (1.8047)	2.4407 (2.2908)	1.0088 (1.6050)	2.4086 (2.2275)

5.2 Objective 2 – SDQ Cut-off for Screening Purpose

5.2.1 Pre-existing Cut-offs

Based on the existing cut-off levels outlined by previous studies, children can be classified into different groups, and then the proportions of each group of individuals in relation to the whole sample were calculated. Table 5-3 and Table 5-4 show the predicted prevalence of mental status for the children based on their post-PAX SDQ scores. General agreement was achieved that about 80% of children are considered mentally healthy, which was close to the statement that one in five Canadian children are suffering from mental health problems and illnesses.

Table 5-3. Prevalence of Mental Disorders Determined by **Goodman's** Cut-offs Based on Post-PAX SDQ

	Abnormal
Total difficulty score	23.7%
Emotional problem score	9.6%
Conduct problem score	18.2%
Hyperactivity score	23.8%
Peer problems score	12.7%

Table 5-4. Gender-Specific Prevalence of Mental Disorders Determined by the **Danish** Cut-offs Based on Post-PAX SDQ

		Abnormal
Boy (N=986)	**Total difficulty score**	23.20%
	Emotional problem score	14.10%
	Conduct problem score	17.00%
	Hyperactivity score	21.80%
	Peer problems score	24.80%
Girl (N=1,007)	**Total difficulty score**	26.90%
	Emotional problem score	15.20%
	Conduct problem score	20.90%
	Hyperactivity score	30.40%
	Peer problems score	17.50%

To draw statistical conclusions on how well the developed classification methods identify students with mental disorders in Manitoba, several validation measures as mentioned in Section 0 were calculated and provided in Table 5-5 and Table 5-6. It is noticed that though the accuracy achieved by Goodman's and Danish cut-offs is relatively high, the sensitivity statistics are unacceptable because our fundamental goal is to screen students with potential mental disorders and prevent delayed, severe outcomes.

Table 5-5. The Goodness of Goodman's cut-offs

		Cut-off	Accuracy	Sensitivity (Se)	Specificity (Sp)	Se+Sp	Distance	Cost
Unisex	Total difficulty	11	0.774	0.541	0.811	1.352	0.497	596
	Emotional problem	4	0.822	0.200	0.920	1.120	0.804	408
	Conduct problem	2	0.799	0.430	0.857	1.286	0.588	517
	Hyperactivity	5	0.778	0.559	0.813	1.372	0.479	593
	Peer problems	3	0.812	0.278	0.896	1.174	0.730	449

Table 5-6. The Goodness of Danish cut-offs

		Cut-off	Accuracy	Sensitivity (Se)	Specificity (Sp)	Se+Sp	Distance	Cost
Unisex	Total difficulty	11	0.774	0.541	0.811	1.352	0.497	596
	Emotional problem	3	0.791	0.270	0.873	1.143	0.741	489
	Conduct problem	2	0.799	0.430	0.857	1.286	0.588	517
	Hyperactivity	5	0.778	0.559	0.813	1.372	0.479	593
	Peer problems	2	0.771	0.433	0.824	1.257	0.594	574
Boy	Total difficulty	13	0.769	0.503	0.827	1.330	0.527	317
	Emotional problem	3	0.765	0.237	0.880	1.117	0.772	274
	Conduct problem	3	0.778	0.356	0.870	1.226	0.657	282
	Hyperactivity	7	0.779	0.492	0.842	1.333	0.533	305
	Peer problems	2	0.736	0.458	0.797	1.255	0.579	341
Girl	Total difficulty	8	0.762	0.667	0.771	1.438	0.404	302
	Emotional problem	3	0.817	0.333	0.867	1.200	0.680	215
	Conduct problem	1	0.781	0.441	0.815	1.256	0.589	262
	Hyperactivity	3	0.723	0.645	0.731	1.376	0.445	339
	Peer problems	2	0.804	0.387	0.847	1.234	0.632	233

5.2.2 Our Proposed Cut-offs

As a result, an exhaustive listing approach was applied to find a set of better classification values. Accuracy, sensitivity, specificity, the distance to the top-left corner of the ROC plane, and healthcare cost are computed for each cut-off threshold. Figure 5-3 is a summary of all ROC curves for every cut-off identified by the SDQ scores/subscales with different genders.

Figure 5-3. ROC of All Cut-offs for Total and Individual Difficulty Subscale Scores by Gender.

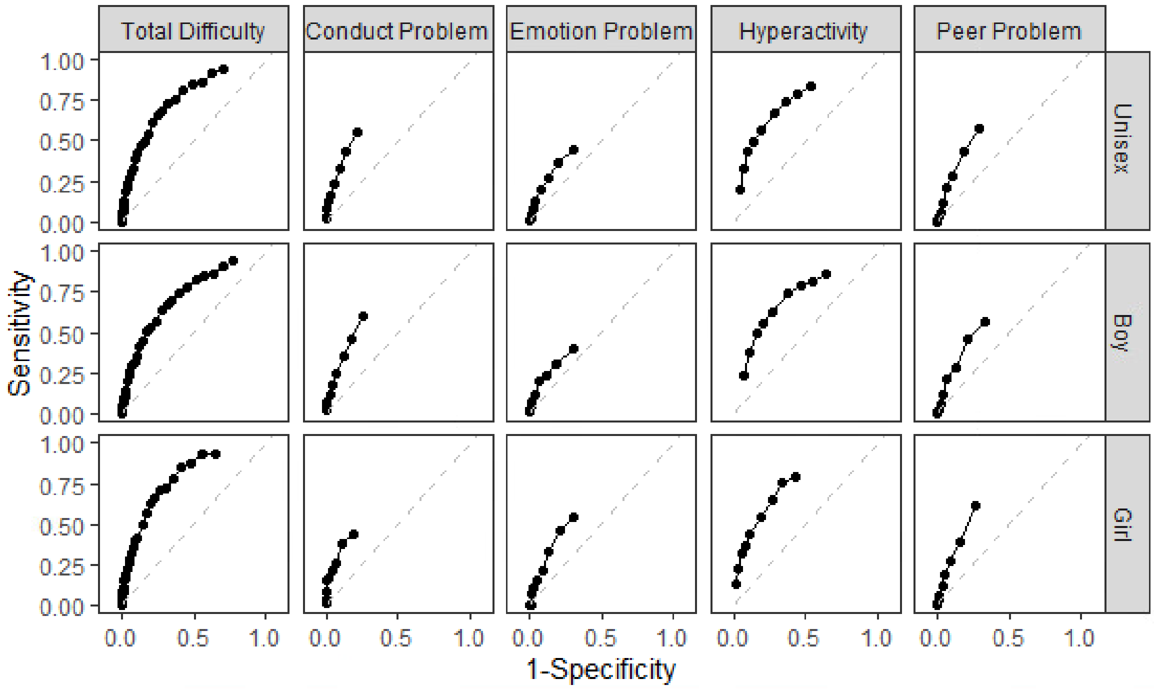
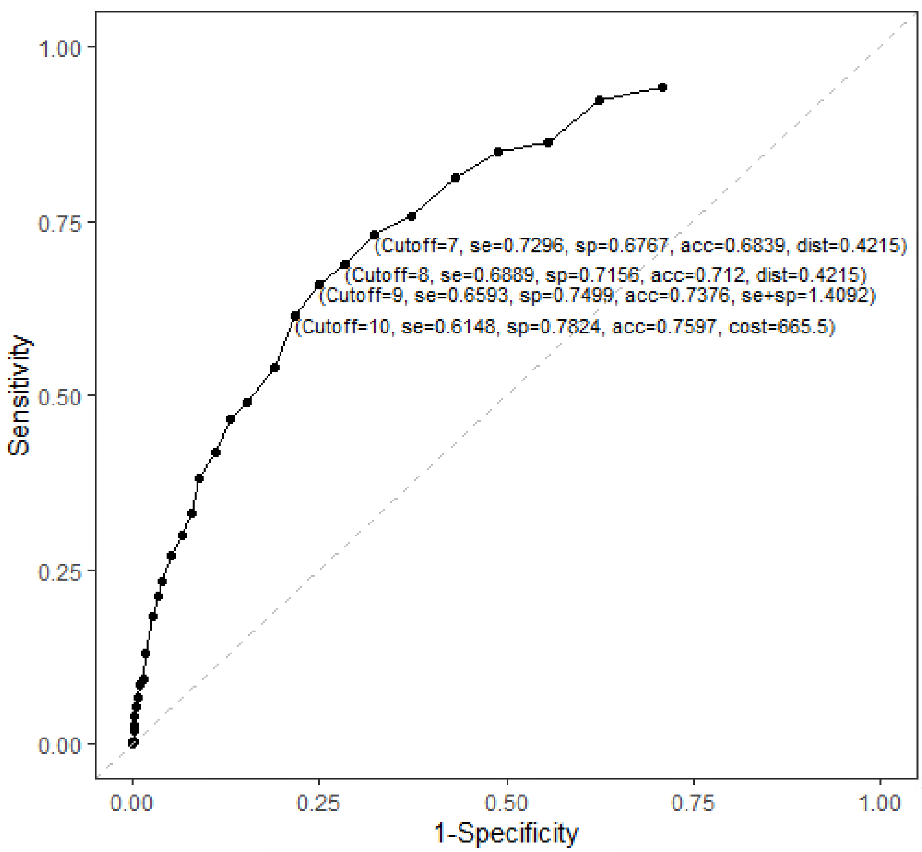


Figure 5-4. ROC of the SDQ Total Difficulty Score



To classify students with the most comprehensive SDQ summary, the SDQ total difficulty was used as one measure to answer the question with multiple sub-domains. Figure 5-4 exhibits all possible cut-offs to classify students using total difficulties and highlights the four "best" points by using different optimization approaches.

Table 5-7, Table 5-8 and Table 5-9 provide all numerical details to our optimal cut-off identified in every SDQ scale using the three aforementioned optimization methods (maximal sum of sensitivity and specificity, minimal distance, and minimal health cost). In general, the total difficulty score and hyperactivity subscale showed high sensitivity in identifying mental disorders, and our approach also demonstrated needs for gender-specific classification. Comparing with the previously developed cut-offs, the proposed values give higher sensitivity to detect the students truly having mental disorders with relatively high accuracy.

Table 5-7. Optimal Cut-offs with Maximal Sum of Sensitivity and Specificity

		Cut-off	**Accuracy**	**Sensitivity (Se)**	**Specificity (Sp)**	**Se+Sp**	**Distance**	**Cost**
Unisex	Total difficulty	9	0.738	0.659	0.750	1.409	0.423	669.5
	Emotional problem	2	0.738	0.359	0.797	1.157	0.672	790.5
	Conduct problem	1	0.748	0.548	0.780	1.328	0.503	704
	Hyperactivity	4	0.714	0.674	0.720	1.394	0.430	687
	Peer problems	1	0.689	0.582	0.706	1.287	0.512	749.5
Boy	Total difficulty	9	0.686	0.678	0.687	1.365	0.449	417.5
	Emotional problem	4	0.797	0.192	0.930	1.122	0.811	491.5
	Conduct problem	1	0.715	0.605	0.739	1.344	0.474	422.5
	Hyperactivity	4	0.642	0.746	0.619	1.365	0.458	421
	Peer problems	2	0.736	0.458	0.797	1.255	0.579	451
Girl	Total difficulty	7	0.736	0.710	0.739	1.448	0.391	266.5
	Emotional problem	2	0.757	0.462	0.787	1.249	0.578	290.5
	Conduct problem	2	0.843	0.376	0.891	1.267	0.633	259
	Hyperactivity	2	0.671	0.753	0.663	1.416	0.418	293
	Peer problems	1	0.725	0.613	0.736	1.349	0.468	285.5

Table 5-8. Optimal Cut-offs with Minimal Distance from the Top-left Corner of ROC curve

		Cut-off	Accuracy	Sensitivity (Se)	Specificity (Sp)	Se+Sp	Distance	Cost
Unisex	Total difficulty	7	0.684	0.730	0.677	1.406	0.422	694.5
	Emotional problem	1	0.663	0.444	0.698	1.142	0.633	830.5
	Conduct problem	1	0.748	0.548	0.780	1.328	0.503	704
	Hyperactivity	4	0.714	0.674	0.720	1.394	0.430	687
	Peer problems	1	0.689	0.582	0.706	1.287	0.512	749.5
Boy	Total difficulty	9	0.686	0.678	0.687	1.365	0.449	417.5
	Emotional problem	1	0.643	0.396	0.697	1.093	0.676	513.5
	Conduct problem	1	0.715	0.605	0.739	1.344	0.474	422.5
	Hyperactivity	4	0.642	0.746	0.619	1.365	0.458	421
	Peer problems	1	0.652	0.565	0.671	1.236	0.545	464
Girl	Total difficulty	7	0.736	0.710	0.739	1.448	0.391	266.5
	Emotional problem	1	0.683	0.538	0.698	1.236	0.552	317
	Conduct problem	1	0.781	0.441	0.815	1.256	0.589	281.5
	Hyperactivity	2	0.671	0.753	0.663	1.416	0.418	293
	Peer problems	1	0.725	0.613	0.736	1.349	0.468	285.5

Unfortunately, for the third optimization approach, no healthcare-cost-related information for mental assessments and treatments was found in the literature to the best of my knowledge. But, we are aware that the early treatment fees are lower than the delayed treatment fees (Donovan & Spence, 2000; Kutcher et al., 2010), and the treatment costs are supposed to be higher than the assessment costs. Therefore, we made a naïve assumption of the cost ratio for $\text{early treatment fee: delayed treatment fee: clinical assessment cost}$ at $1:3:0.5$. Then, the optimal cut-offs were found as follows. Notice that the sensitivity measures achieved by this method, at least with the hypothetical cost, are generally not high as desired (see the table below). Subsequently, these findings are only reported as references for discussion and future research.

Table 5-9. Optimal Cut-offs achieved by Minimal Healthcare Cost

		Cut-off	Accuracy	Sensitivity (Se)	Specificity (Sp)	Se+Sp	Distance	Cost

Unisex	Total difficulty	10	0.760	0.615	0.782	1.397	0.442	665.5
	Emotional problem	4	0.822	0.200	0.920	1.120	0.804	771
	Conduct problem	2	0.799	0.430	0.857	1.286	0.588	701.5
	Hyperactivity	7	0.835	0.433	0.898	1.331	0.576	664
	Peer problems	2	0.771	0.433	0.824	1.257	0.594	728
Boy	Total difficulty	10	0.710	0.638	0.726	1.364	0.454	416
	Emotional problem	4	0.797	0.192	0.930	1.122	0.811	491.5
	Conduct problem	1	0.715	0.605	0.739	1.344	0.474	422.5
	Hyperactivity	6	0.753	0.554	0.796	1.350	0.491	417.5
	Peer problems	2	0.736	0.458	0.797	1.255	0.579	451
Girl	Total difficulty	13	0.860	0.398	0.907	1.305	0.609	247.5
	Emotional problem	6	0.896	0.108	0.976	1.083	0.893	270
	Conduct problem	6	0.914	0.151	0.991	1.142	0.850	255
	Hyperactivity	7	0.890	0.323	0.948	1.270	0.679	243
	Peer problems	3	0.856	0.269	0.916	1.185	0.736	267.5

In summary, as our primary goal is to identify as many students who actually have mental disorders as possible, the results will be finalized by filtering out the cut-off values having the highest sensitivity from the optimal classification values presented by all three methods. Therefore, we present the following table with our proposed cut-offs specified by gender for screening purposes.

Table 5-10. Proposed, Gender-Specific Cut-off

	Unisex		**Boy**		**Girl**	
	Normal	Abnormal	Normal	Abnormal	Normal	Abnormal
Total difficulty	0-7	8-40	0-9	10-40	0-7	8-40
Emotional problem	0-1	2-10	0-1	2-10	0-1	2-10
Conduct problem	0-1	2-10	0-1	2-10	0-1	2-10
Hyperactivity	0-4	5-10	0-4	5-10	0-2	3-10
Peer problems	0-1	2-10	0-1	2-10	0-1	2-10

5.3 Objective 3 – Use of SDQ for Screening

5.3.1 Mental Disorder Prediction with SDQ

Although the classification using the SDQ total difficulty score allows quick and straightforward screening of mental disorders, we are also interested to know the possibility to improve the classification by considering the effects of the SDQ impact scale (Q26) in addition to merely SDQ total difficulty score. Here, only Q26 from the impact-related questions was evaluated due to the massive amount of missing information in the other impact-scaled items. Out of 4,676 observations in our study cohort, only less than 800 students have complete impact scores for Q27–30 due to the nature that the completion of Q27–30 is conditioning on the impactful difficulty level answered on Q26 (see the second page of SDQ; Appendix C). If we carried out the analysis with all five impact items, we would not only suffer from an unneglectable amount of information loss together with other incompleteness from Q1–25 but also generate misleading conclusions. Thus, only Q26 was used as an impact covariate for advancing the classification, and our sample size for analysis remained at 1,993.

Naturally, we proceeded with the analyses using series of mixed-effect logistic regression models accounting for the structure that the students are nested in classes. Model A-1 predicted students' mental health status using only the SDQ total difficulty scores (the model composition is provided below). To improve upon the previous analysis, Model A-2 incorporated the impact scale in addition to the SDQ total difficulty scores, and Model A-3 added another layer of discussion if the interaction between SDQ total difficulty and impact scales has a significant impact on the students' mental outcome prediction. To validate this part of the analyses, the dataset was splitting into training and testing sets, where data was modeled and analyzed on the training set to generate the model estimates, and then its goodness of cut was validated on the testing set with the sensitivity, specificity, distance and cost measures, and ROC curves.

Model A-1. Predicting Mental Disorder with Total Difficulty

Level 1: Inter-personal variability within a class

$$\ln\left(\frac{p}{1-p}\right) = \beta_{0j} + \beta_{1j} \cdot W_{ij}$$

Level 2: Variability among classes (random intercept)

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Here, p denotes the probability of the student with mental disorder(s); W_{ij} is the total difficulty score of the student i in the classroom j; β_{0j} and β_{1j} are level 1 model parameters for intercept and slope; γ_{00} is the level 2 model parameter for intercept; u_{0j} is the random class effect.

On top of Model A-1, Model A-2 only added the covariate of the impact scale at the level 1 sub-model, and Model A-3 simply included additional impact and the interaction between total difficulty score and impact at the level 1 sub-model; hence, detailed mathematical representations for these two models are omitted. Visually speaking, the green (Model A-2) and blue (Model A-3) curves are generally above the red curve (Model A-1), meaning better model performances were obtained by the two models with more information. Table 5-11 displays the estimation results, fit statistics as well as odds ratio summary for all three models, and this table restates the same story. Even Model A-1 has demonstrated nearly excellent discrimination (AUC=78.99%) according to Hosmer (2013), Model A-2 and Model A-3 advance the predicting power (AUC>80%; i.e. excellent discrimination according to Hosmer (2013)) by including the impact scale and/or an interaction term considering both the SDQ difficulty score and the impact score. Therefore, students' mental outcome classification analysis can be further examined with extra SDQ components for better prediction.

Figure 5-5. ROC Curves for Model A-1 to Model A-3 using Testing Set

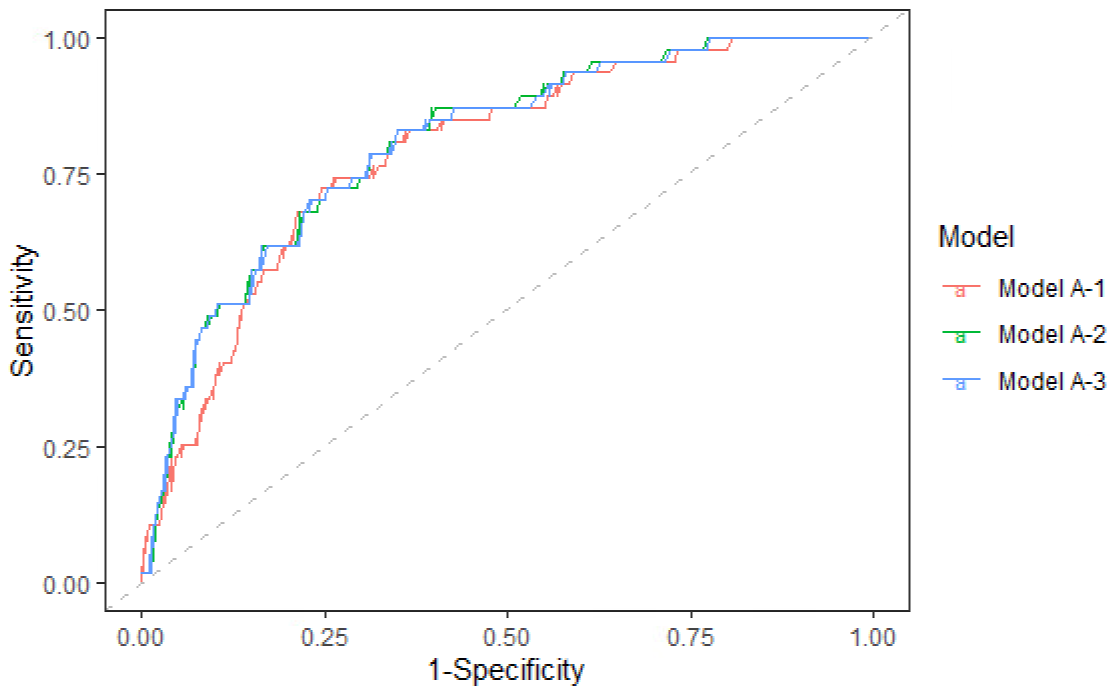


Table 5-11. Results of Model A-1 to Model A-3 using Training Set

	Model A-1 Estimation (SD)			Model A-2 Estimation (SD)			Model A-3 Estimation (SD)		
Fixed Effect									
Intercept	-3.1330	(0.1708)	***	-3.0767	(0.1685)	***	-3.1299	(0.1958)	***
Total difficulty	0.1388	(0.0121)	***	0.0680	(0.0162)	***	0.0754	(0.0210)	***
Impact				0.7114	(0.1203)	***	0.7904	(0.1867)	***
Total difficulty*Impact							-0.0065	(0.0118)	
Random Effect (Between-class)									
Intercept	0.1010			0.0938			0.0928		
Fit Statistics									
AUC	78.99%			80.66%			80.55%		
Deviance	1101.9			1064.9			1064.6		
AIC	1107.9			1072.9			1074.6		
BIC	1124.0			1094.4			1101.5		
Odds Ratio	OR	95% CI		OR	95% CI		OR	95% CI	
Abnormal	1.15	(1.12, 1.18) *		1.07	(1.04, 1.1) *		N.A.	N.A.	
Impact				2.04	(1.61, 2.58) *		N.A.	N.A.	

Note: p-value <0.1 +, p-value <0.05 *, p-value <0.01 **, p-value<0.001 ***; the p-values of the random effects are not provided; * next to the CI denotes the comparisons that showed significant difference.

In short, Model A-1 shows that those students with higher SDQ difficulty scores are more likely to have mental disorders. Every one-unit higher in the SDQ difficulty score is associated with a 14.9% (95% CI: 12%–18%) increase in odds of mental disorder. Model A-2 indicates those with higher SDQ impact scores are more likely to have mental disorders, and Model A-3 indicates that the interaction effect of the SDQ difficulty score and impact score is not statistically significant.

Instead of treating SDQ total difficulty score as a continuous variable, we also examined the compatibility between mental outcome and the abnormal mental health classified with our cut-off proposed in Section 0. Besides, we are also interested in if adding impact scales and an interaction effect can advance our prediction. Similar to the previous model building steps, Model A-4 started with the classified SDQ category as the only predictor for the mental outcomes of students. Gradually, the impact scale and the interaction between the SDQ classification and impact are added in Model A-5 and Model A-6 respectively. The composition of Model A-4 is provided below and similar to the composition of Model A-5 and Model A-6 except for the presence of extra predictors in the two latter models. Table 5-12 and Figure 5-6 summarize the comparisons for these three models predicted with the proposed cut-offs.

Model A-4. Predicting Mental Disorders with the Hyperactivity Subscale

Level 1: Inter-personal variability within a class

$$\ln\left(\frac{p}{1-p}\right) = \beta_{0j} + \beta_{1j} \cdot W'_{ij}$$

Level 2: Variability among classes (random intercept)

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Here, W'_{ij} denote the mental disorder classification of student i in the classroom j using total difficulty score with our proposed cut-off, and the notations for other parameters stay the same as before.

Figure 5-6. ROC Curves for Model A-4 to Model A-6 using Testing Set

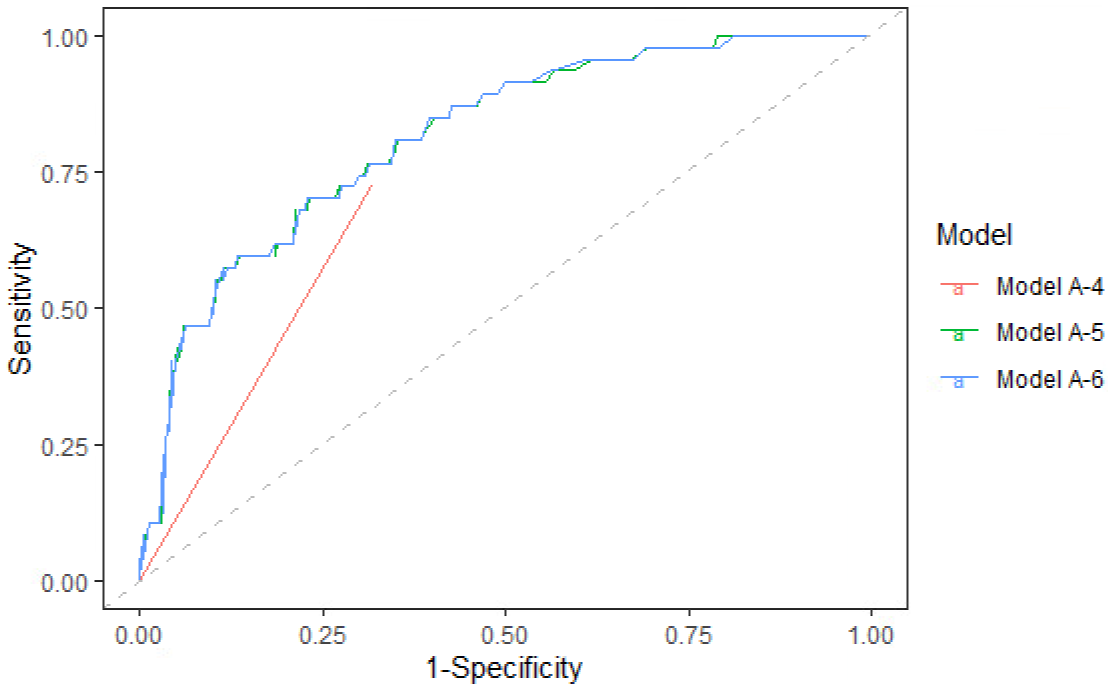


Table 5-12. Results of Model A-4 to Model A-6 using Training Set

	Model A-4 Estimation (SD)			Model A-5 Estimation (SD)			Model A-6 Estimation (SD)		
Fixed Effect									
Intercept	-2.7289	(0.1332)	***	-2.8985	(0.1531)	***	-2.9438	(0.1664)	***
Abnormal	1.7291	(0.1617)	***	0.5830	(0.2225)	**	0.6791	(0.2545)	**
Impact				0.8965	(0.1103)	***	1.0906	(0.2743)	***
Abnormal * Impact							-0.2257	(0.2935)	
Random Effect (Between-class)									
Intercept	0.0000			0.0529			0.0531		
Fit Statistics									
AUC	70.38%			81.32%			81.35%		
Deviance	1157.0			1076.0			1075.5		
AIC	1163.0			1084.0			1085.5		
BIC	1179.1			1105.5			1112.3		
Odds Ratio	OR	95% CI		OR	95% CI		OR	95% CI	
Abnormal	5.64	(4.10, 7.74) *		1.79	(1.16, 2.77) *		N.A.	N.A.	
Impact				2.45	(1.97, 3.04) *		N.A.	N.A.	

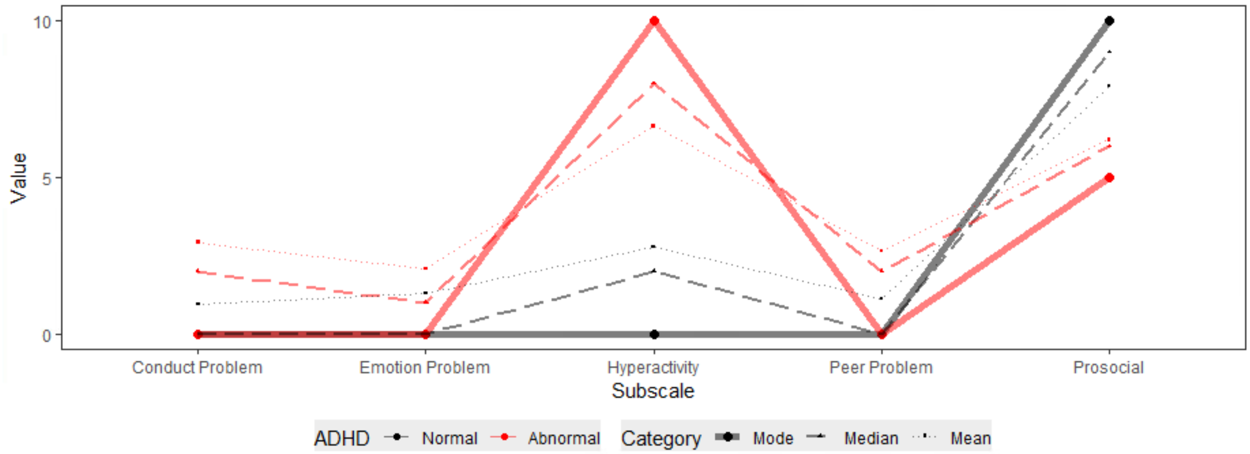
Note: p-value<0.1 +, p-value<0.05 *, p-value<0.01 **, p-value<0.001 ***; the p-values of the random effects are not provided; * next to the CI denotes the comparisons that showed significant difference.

Both Figure 5-6 and Table 5-12 display that including impact scale advances the prediction. Model A-4 suggests that the students identified as mentally abnormal have 4.64 (95% CI: 3.10-6.74) times increased odds of having mental disorders compared to those who identified as normal by our cut-off. From Model A-5, a unit increase in the SDQ impact score is associated with 2.45 times odds of having mental disorders. Model A-6 indicates that the interaction effect of the classified mental outcome and impact score is not statistically significant, which is similar to the summary of Model A-3.

5.3.2 ADHD Diagnosis Prediction with Logistic Models

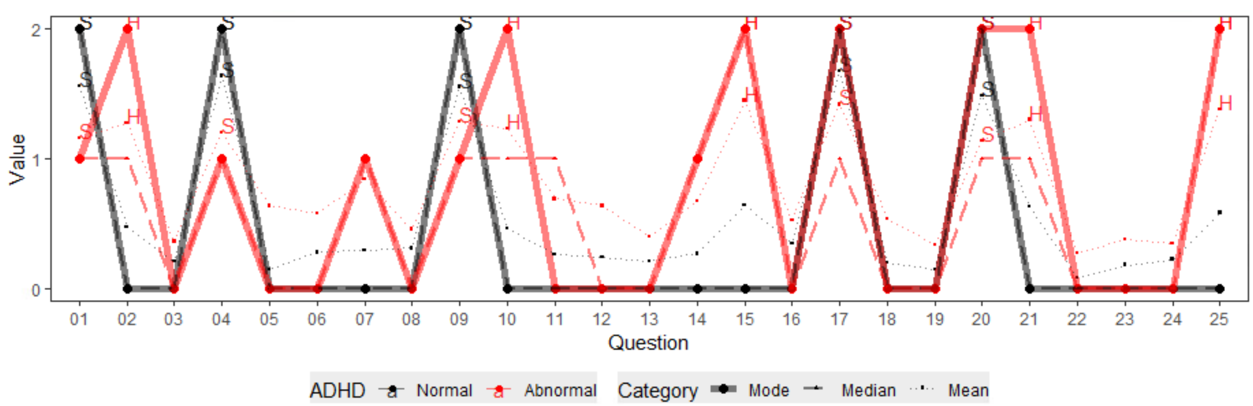
As there is high concordance in the ADHD case definition and the description of the SDQ questions in the hyperactivity sub-domain, we plotted Figure 5-7 and Figure 5-8 to verify whether it is reasonable to assume that the students' disease presence with ADHD can be classified using SDQ subscale and itemized scores. Preliminarily, Figure 5-7 and Figure 5-8 demonstrate the distributional difference between the students with or without ADHD. It is noticed that the hyperactivity subscale scores are extremely high for students with ADHD diagnoses (represented by the red curves) while the scores for the conduct problem, emotional problem, hyperactivity and peer relationship problem subscales are obvious on the lower end. As a result, we may examine the discriminative property of the hyperactivity subscale to detect this particular disease.

Figure 5-7. Distribution of the SDQ Subscales of Students with/without an ADHD diagnosis



When looking at the itemized distributional plot, the values of statistics over 1 are signaled with a letter showing which sub-domain this item belongs to. Figure 5-8 reveals the different patterns of the SDQ item scores for two groups of students. This further confirmed the assumption that the hyperactivity subscale could be a good measure to tell whether a student is more prone to have ADHD since most of the differences in the curve patterns happen at the hyperactivity-related items.

Figure 5-8. Distribution of the SDQ Items of Students with/without an ADHD diagnosis



Therefore, we examined the predictive capacity of the hyperactivity subscale to the ADHD presences. First, Model B-1 was constructed as an initial-evaluation or reference model predicting ADHD presence using only hyperactivity subscale score (mathematical representation provided below). Model B-2 overviewed the possible improvement in prediction by including the impact supplement, and Model B-3 uncovered the possible improvement by including hyperactivity subscale score, impact score, and their interaction effect.

Model B-1. Predicting ADHD Presences with the Hyperactivity Subscale

Level 1: Inter-personal variability within a class

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot Z_{ij}$$

Level 2: Variability among classes (random intercept)

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Z_{ij} denotes the hyperactivity subscale's score of student i in the classroom j, and the notations for other parameters stay the same as before.

Figure 5-9. ROC Curves for Model B-1 to Model B-3 using Testing Set

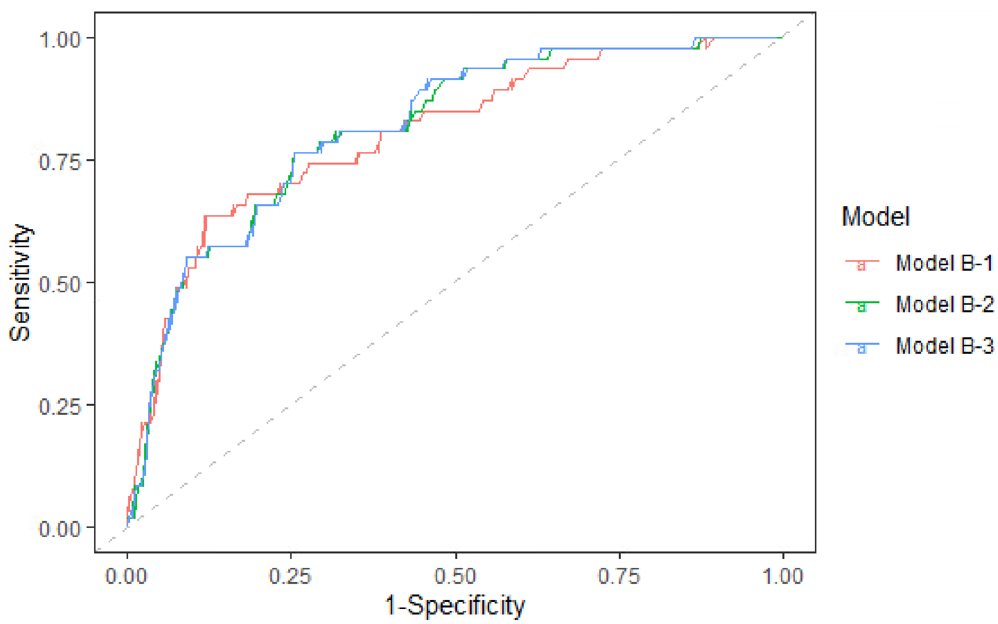


Table 5-13. Results of Model B-1 to Model B-3 using Training Set

	Model B-1 Estimation (SD)			Model B-2 Estimation (SD)			Model B-3 Estimation (SD)		
Fixed Effect									
Intercept	-3.8074	(0.2173)	***	-3.8301	(0.2211)	***	-3.7749	(0.0023)	***
Hyperactivity	0.3518	(0.0300)	***	0.1794	(0.0378)	***	0.1631	(0.0023)	***
Impact				0.8622	(0.1273)	***	0.7582	(0.0023)	***
Hyperactivity* Impact							0.0175	(0.0022)	***
Random Effect (Between-class)									
Intercept	0.1758			0.2328			0.2433		
Fit Statistics									
AUC	84.16%			84.92%			84.93%		
Deviance	896.8			846.0			845.7		
AIC	902.8			854.0			855.7		
BIC	918.9			875.5			882.6		
Odds Ratio	OR	95% CI		OR	95% CI		OR	95% CI	
Hyperactivity	1.42	(1.34, 1.51) *		1.20	(1.11, 1.29) *		N.A.	N.A.	
Impact				2.37	(1.85, 3.04) *		N.A.	N.A.	

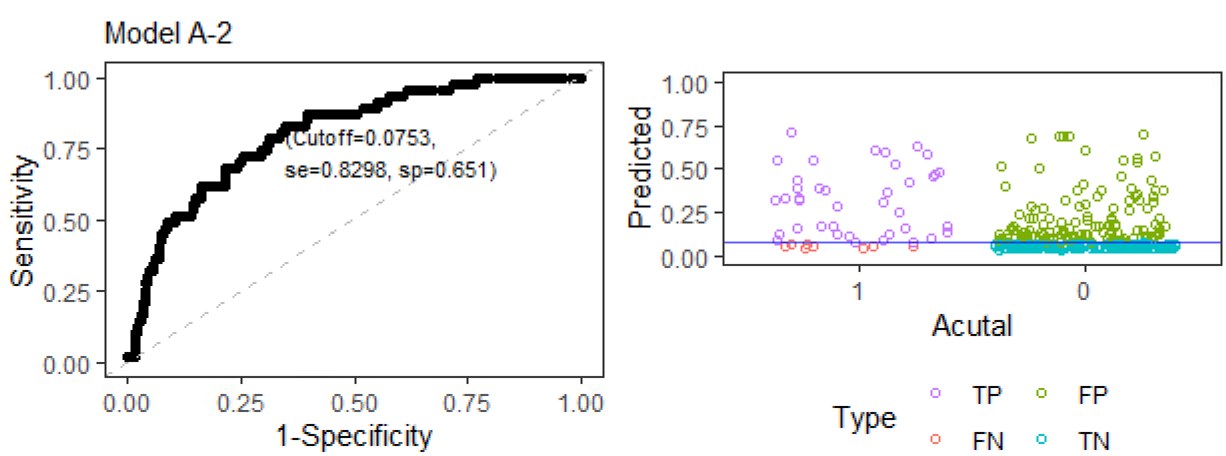
Note: p-value <0.1 +, p-value <0.05 *, p-value <0.01 **, p-value<0.001 ***; the p-values of the random effects are not provided; * next to the CI denotes the comparisons that showed significant difference.

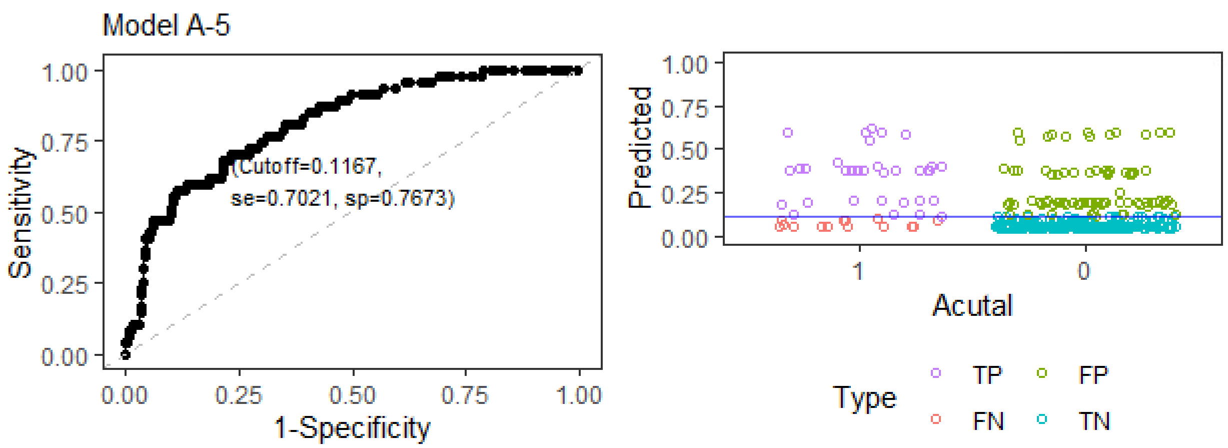
From Figure 5-9 and Table 5-13, hyperactivity is a good measure to predict ADHD status. Incorporating the impact scale slightly advances the predictive power. Model B-1 demonstrated that every unit increase in the hyperactivity score results in 42% (95% CI: 34%–51%) increased odds of discovering ADHD in students. Model B-2 showed that impact scale is useful towards students' ADHD presence prediction, and Model B-3 exhibits that not only impact scale is important for prediction, but its interaction effect with the hyperactivity subscale is also significant.

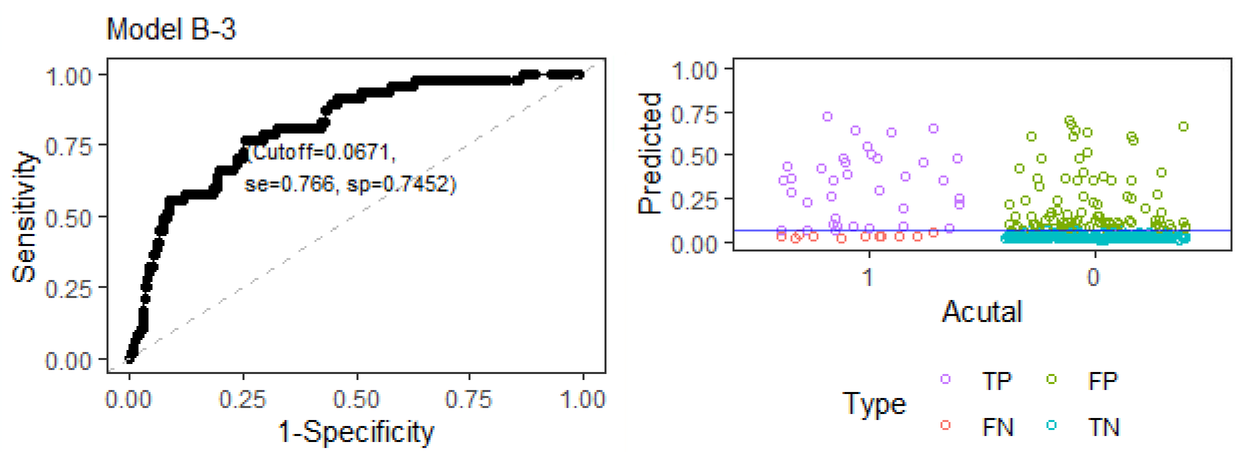
Compared to the previous models, Model B-series is particularly substantial as it has a simple construct with excellent prediction and focuses only on the ADHD disease group. By using these models, it allows high efficiency by reducing the pre-referral screening time to direct this group of children to the ADHD-related specialists.

The cut-off can also be developed using the selected models from above, Model A-2, Model A-5 and Model B-3 stood out from the parallel comparisons as their fit statistics are most desirable. In specific, Model A-2 has the highest AUC (80.66%) and lowest Deviance (1064.9), AIC (1072.9), as well as BIC (1094.4) values among the first three comparable models. Model A-5 has reasonably high AUC (81.32 %) and low Deviance (1076.0), AIC (1084.0), as well as BIC (1105.5) values among the three comparable models. Model B-3 has the highest AUC (84.93%) and reasonably low Deviance (845.7), AIC (855.7), as well as BIC (882.6) values among the three comparable models. The cut-off selection process was carried out similarly in the previous subsection, and the following criteria were used: first, one of the following conditions is met: 1) maximal sum of sensitivity and specificity, 2) the minimal distance to the top-left corner of the ROC plane, or 3) the minimal cost of the total hypothesized healthcare fee is satisfied; second, the highest sensitivity is achieved. In Figure 5-10, the optimal probability cut-off found for each model is labeled in the ROC curves (on the left-hand side), and the visual presentation of classified students with selected probability cut-off is provided by the confusion matrix presentation on the right. Here, the purple and turquoise colored points represent the correct-specified, majority of the observations.

Figure 5-10. Classification by Model A-2, Model A-5, and Model B-3 using Testing Set







In the last two subsections, different approaches are outlined to classify students' mental health. Section 0 focused on using only the SDQ total difficulty scores for quick and convenient detection of mental outcomes, where no model is required for classification (validated using the Gold Standard). Whereas, this section screened students using different predictors in the logistic model that gives us more information on the students' mental health or the possibility of having ADHD (validated with training and testing sets). In summary, the SDQ scale score can help to predict mental disorders. The combination of SDQ scales and SDQ impact score can improve prediction. The SDQ subscale such as hyperactivity score can help to predict a specific mental disorder, ADHD. The combination of SDQ subscales and impact scale can increase prediction accuracy. Finally, Model A-2, Model A-5 and Model B-3 outperformed the comparable models, which not only showed the best prediction, but also gave the cut-off probabilities for more precise classification. The probability-based cut-offs are listed as follows. If the predicted probability is over the listed cut-off, then one can classify this student with mental or ADHD outcomes.

Table 5-14. Cut-offs with Multiple SDQ Components

Predicted Outcome	**Mental Disorder**		**ADHD Presence**
Predictive Model Components	Total difficulty + Impact	Abnormal + Impact	Hyperactivity Subscale + Impact + Interaction
Cut-off Probability	0.0753	0.1167	0.0671

Note: The cut-offs are selected by two criteria: 1) Either max(se+sp), min distance to the top left corner, or min cost with hypothesized cost with each classification is satisfied, 2) Highest sensitivity.

5.4 Objective 4 – Implication of the PAX Program Evaluation

Different individual trajectories are shown for most students, but there exists a general decreasing trend for the SDQ total difficulty scores in students involved in the PAX program. Explicitly, one can see the PAX students generally started with a higher SDQ total difficulty score, and their difficulty scores dipped right after students received the PAX training (see Figure 5-11; the black curves represent individual trajectories, and the red curve represents the summarized trajectory by the median of groups). Figure 5-12 also addresses the same information from a different perspective, a bigger proportion of students in the PAX group started with abnormal mental health, and this proportion decreased after they received the PAX program (the black curves represent individual trajectories of having or not having abnormal mental health, and the red curve represents the proportion of students having abnormal mental health). In comparison, the students in the control group showed almost no changes in the proportion of students with abnormal mental health throughout the study period. While the trends appear promising, rigorous effectiveness examination with statistical tests is required to draw conclusions whether there exists a statistically significant decline in the PAX group.

Figure 5-11. Longitudinal Trend of SDQ Total Difficulty Score by the Exposure to the PAX program

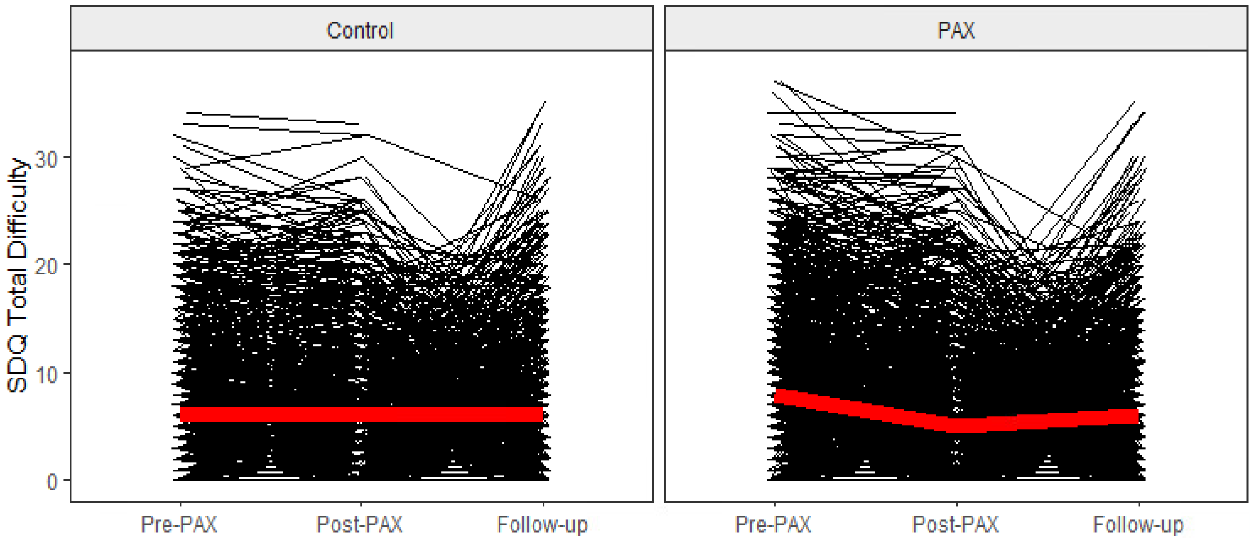
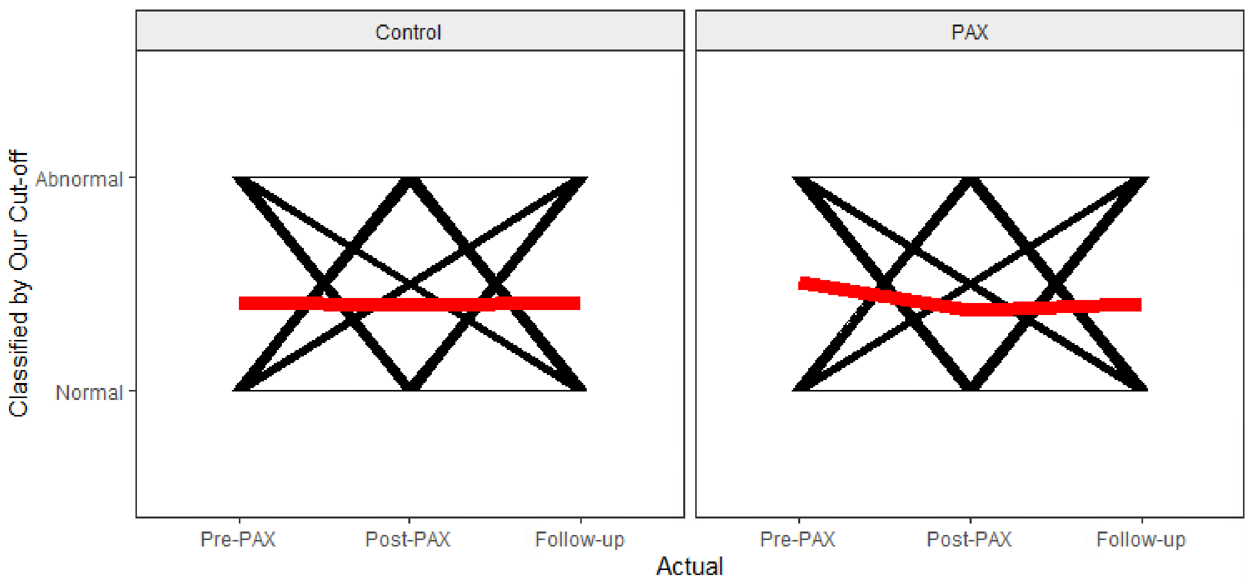


Figure 5-12. Longitudinal Trend of SDQ Classification by the Exposure to the PAX program



5.4.1 The PAX Program Evaluation In Terms of the Risk of Abnormal

We are interested in finding whether the PAX program was effective in reducing the likelihood of students having abnormal mental health. Here, we leveraged the study findings from the previous subsections, and classified students' mental health with the proposed cut-offs, 7. Comparisons were conducted in contrast to the classification result with Goodman's and Danish cut-off at 11. In the table below, a preparatory prevalence summary is provided for both cut-offs. In brief, both approaches show the prevalence of students having mental outcomes decreased right after they receive the PAX program, and the probability slightly increased at the grade five follow-up.

Table 5-15. Longitudinal Summary of Prevalence of Mental Disorder using Goodman's vs. Our Proposed Cut-off

	Pre-existing Cut-off=11	Our Proposed Cut-off=7
Pre-PAX	29.9%	46.9%
Post-PAX	23.5%	38.1%
Follow-up	26.2%	41.1%

Longitudinal models of three levels were constructed as described in equations (5(7) to account for the cluster-sampling nature of the study data. In the first level of the model, it quantifies the intra-personal change over time, the second level model quantifies how the inter-personal variability affects the SDQ scores of students, and the third level model quantifies how students' SDQ scores change among different classes. Model C-1 is an unadjusted model to predict the effect of the PAX exposure and time on the probability of students being classified as abnormal with the Goodman's and Danish cut-off (see the mathematical representation below), while Model C-2 has the same composition as Model C-1 except for a different outcome of interest, the probability defined using the proposed cut-off in this paper.

Model C-1. Unadjusted model, SDQ Total Difficulty Cut-off at 11 (Goodman's and Danish Cut-off)

Level 1: Intra-personal change over time (from pre-PAX to G5)

$$\ln\left(\frac{p}{1-p}\right) = \pi_{0ij} + \pi_{1ij} \cdot time$$

Level 2: Inter-personal variability within a class (random intercept)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j}$$

Level 3: Variability among classes (random intercept)

$$\beta_{00j} = \gamma_{000} + \gamma_{001} \cdot \text{PAX} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + \gamma_{101} \cdot \text{PAX}$$

Model estimations and odds ratio are compared in Table 5-16 for the mental outcome classified using Goodman's cut-off versus our proposed cut-off for SDQ total difficulty score.

Table 5-16. Comparison for Model Estimation using Pre-existing Cut-off (at 11) vs. Our Proposed Cut-off (at 7)

	Model C-1 Estimation (SD)			**Model C-2** Estimation (SD)		
Fixed Effects						
Intercept	-1.3308	(0.1403)		-0.3738	(0.1245)	
Time	0.0038	(0.0538)		-0.0058	(0.0475)	
PAX	0.4815	(0.1790)	**	0.5456	(0.1631)	***
Time*PAX	-0.1694	(0.0718)	*	-0.2166	(0.0638)	***
Random Effects						
At Level 2 (Student):						
Intercept	0.9163	(0.1258)		0.6340	(0.0904)	
At Level 3 (Class):						
Intercept	0.4220	(0.0674)		0.4168	(0.0603)	
Odds Ratio	OR	95% CI		OR	95% CI	
Change (PAX group)	0.847	(0.772, 0.930)	*	0.801	(0.737, 0.870)	*
Change (Control group)	1.004	(0.903, 1.115)		0.994	(0.906, 1.091)	

Note: p-value<0.1 +, p-value<0.05 *, p-value<0.01 **, p-value<0.001 ***; the p-values of the random effects are not provided

Table 5-16 shows that no significant change in the likelihood of being abnormal mental health was found for the control group, but a significant decrease was found for the PAX group. Using different cutoffs for 'abnormal' mental health gave us qualitatively the same conclusions. But the effect size in terms of odds of being abnormal based on the two cut-offs might be different.

Results from the random effect part indicate that the variations in the propensity of being abnormal based on our proposed cutoffs are smaller than that based on Goodman's cutoff.

5.4.2 Who Are At High Risk of Abnormal

At the individual model level, several other factors of interest can be investigated for their possible impact on the risk of abnormal mental health. At level 1, the variables of considerations are the status of whether the student is a Child in Care (CIC) in Child and Family Services, the receipt of Voluntary Family Services (VFS), and the receipt of Income Assistance (IA) by the student's family. At level 2, the gender, socioeconomic status (SES) and urbanity of students are considered. Therefore, the next model is constructed to evaluate the longitudinal change of SDQ categories at different levels and the impact of selected demographic factors. Descriptive summary and model comparison are provided in the following sections.

To better understand our study cohort and help to select entrancing factors of control in the models, the demographics and other factors of students are summarized by the exposure to the PAX program in Table 5-17. Most of the tabulated factors are significantly different (or different at certain time points) between groups except for gender and individual student's urban residence status.

Table 5-17. Descriptive summary of demographics and other factors

Variable	Control cohort N=2,100	PAX cohort N=2,576	P-value	
Gender (Male, %)	48.62%	46.78%	0.7880	
Child in Care (CIC) in Child and Family Services (%)				
Pre-PAX	6.95%	8.15%	0.1239	
During-PAX	3.43%	4.11%	0.2225	
Post-PAX	2.29%	3.30%	0.0380	*
Child in Voluntary Family Services (VFS) (%)				
Pre-PAX	14.19%	18.05%	0.0005	***

During-PAX	8.57%	10.68%	0.0158	*
Post-PAX	7.10%	10.21%	0.0002	***
Child in Family Receiving Income Assistance (IA) (%)				
Pre-PAX	22.43%	25.50%	0.0145	*
During-PAX	11.57%	12.66%	0.2595	
Post-PAX	12.95%	14.52%	0.1226	
Student Urbanity (%)	43.14%	41.03%	0.1458	
Student SES (SEFI) (mean, SD)	0.2720 (1.0770)	0.3566 (1.2198)	0.0119	*

Note: p-value<0.1 +, p-value<0.05 *, p-value<0.01 **, p-value<0.001 ***

After including selected predictors from the table above, the Deviance, AIC, and BIC dropped from Model C-2. Some variations in the outcome are explained by the additional predictors such as CIC, VFS, IA and students' urbanity (see Table 5-19). The variations in the tendency of being abnormal for both within classrooms and between classrooms are reduced. Based on the odds ratio summary table provided on the next page, several conclusions can be drawn with the assumption that all predictors other than those being compared are adjusted.

Table 5-18. Fit Statistics Comparing the Unadjusted and Adjusted Models

	Model C-2	Model C-3
Deviance	10488.84	10339.28
AIC	10500.84	1359.28
BIC	10523.12	10396.42

Table 5-19. Model Estimation for Adjusted Model, SDQ Total Difficulty Cut-off at 7 (Our Proposed Cut-off)

	Estimation (SD)		
Fixed Effects			
Intercept	-0.5056	(0.1291)	
Time	0.0397	(0.0475)	
PAX	0.4902	(0.1602)	**
Time*PAX	-0.2056	(0.0637)	**
Child in Care (CIC) in Child and Family Services	0.3608	(0.1574)	*
Child in Voluntary Family Services (VFS)	0.6030	(0.1096)	***
Child in Family Receiving Income Assistance (IA)	0.5537	(0.0886)	***
Student Urbanity	-0.2798	(0.0873)	**

Random Effects		
At Level 2 (Student):		
Intercept	0.5761	(0.0873)
At Level 3 (Class):		
Intercept	0.3581	(0.0544)

Note: p-value<0.1 +, p-value<0.05 *, p-value<0.01 **, p-value<0.001 ***; the p-values of the random effects are not provided

Table 5-20. Odds Ratios of the Selected Model, Cut-off at 7 (Minimal Distance)

	Odds ratio	95% Confidence Interval
Change (PAX Group)	0.847	(0.779 , 0.921) *
Change (Control group)	1.040	(0.948 , 1.142)
Received vs. Not Received CIC	1.435	(1.054 , 1.953) *
Received vs. Not Received VFS	1.828	(1.474 , 2.266) *
Received vs. Not Received IA	1.740	(1.462 , 2.070) *
Urban vs. Rural Students	0.756	(0.637 , 0.897) *

Note: The comparisons are carried out with everything else held the same; * denotes the comparisons that showed significant difference.

Several factors are associated with the likelihood of being in abnormal mental health. Table 5-20 indicates that the odds of being abnormal for students whose family received IA are 1.740 (95% CI: 1.462–2.070) times higher than those who did not. Similarly, those students in CIC and VFS and from rural areas are at high risk of being abnormal. After controlling the effect of these factors, the odds of being in abnormal mental health decreased by 15.3% (95% CI: 7.9%-22.1%) as a result of the PAX program.

5.4.3 Who Have Higher SDQ Difficulty Scores

Further, some interests remain in find out how the additional predictors affect the students' mental wellness and who could have higher SDQ difficulty scores. Due to some loss of statistical power when we categorize the continuous difficulty score to categorical response, the effect of predictors of different levels could not be fully explored with multi-level logistic models. Therefore, we conducted additional analyses by treating the SDQ total difficulty score ranged from 0 to 40 as a continuous variable to tell a more complete story. Three-level models are constructed in a similar manner. Model D-1 checked the significance of considering the factors from a structural perspective (e.g. intra-personal, inter-personal and between-class effect on the SDQ total scores) to a controlling-covariate stand of point (e.g. gender, urbanity, SES, and the service/assistance a student received).

Model D-1. Adjusted Multi-Level Regression Model with Demographic Variables

Level 1: Intra-personal change over time (from pre-PAX to G5)

$$y_{tij} = \pi_{0ij} + \pi_{1ij} \cdot time + \pi_{2ij} \cdot CIC + \pi_{3ij} \cdot VFS + \pi_{4ij} \cdot IA + \varepsilon_{tij}$$

Level 2: Inter-personal variability within a class (random intercept and random slope)

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} \cdot SES + \beta_{12j} \cdot Gender$$
$$\pi_{2ij} = \beta_{20j} + \beta_{21j} \cdot Gender + \beta_{22j} \cdot Urbanity$$
$$\pi_{3ij} = \beta_{30j} + \beta_{31j} \cdot Urbanity + r_{3ij}$$
$$\pi_{4ij} = \beta_{40j} + \beta_{41j} \cdot SES + \beta_{42j} \cdot Urbanity$$

Level 3: Variability among classes (random intercept and random slope)

$$\beta_{00j} = \gamma_{000} + \gamma_{001} \cdot \mathrm{PAX} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + \gamma_{101} \cdot \mathrm{PAX}$$
$$\beta_{20j} = \gamma_{200}$$
$$\beta_{30j} = \gamma_{300} + u_{30j}$$
$$\beta_{40j} = \gamma_{400} + u_{40j}$$

Table 5-21 displays the results of fitting the aforementioned three-level mixed model. In the first section of this table, parameter estimation for each selected fixed effect is provided; followed by random effects estimation and goodness-of-fit for the model. Model D-1 was selected through the model development process as it has the most comprehensive package of answers to the research questions (significance were found in systematic effect and students' gender, urbanity and socioeconomic status) and lowest fit statistics (Deviance statistics: 53077.1, AIC: 53123.1, and BIC: 53208.7). Beyond this point, the random slope effects of the level 2 predictors as well as fixed-effect level 3 predictors were explored, but no significance of these mentioned effects was detected. Therefore, the optimal model specification was achieved by Model D-1 for the SDQ total score considered as a continuous variable.

Table 5-21. Results from Model D-1 for SDQ Total Difficulty Score

	Estimation (SD)		
Fixed Effects			
Intercept	10.2071	(0.3615)	
Time	-0.8064	(0.1379)	***
PAX	1.6053	(0.4396)	***
PAX * Time	-0.6235	(0.1668)	***
Gender (reference = "M")	-3.4107	(0.3603)	***
Time * Gender	0.7772	(0.3600)	***
Student SES (SEFI)	1.1113	(0.2004)	***
Time * Student SES (SEFI)	-0.2915	(0.0826)	***
Child in Care (CIC) in Child and Family Services	0.7062	(0.7247)	
Gender * Child in Care (CIC) in Child and Family Services	1.7701	(0.8168)	*
Student Urbanity	-0.2293	(0.2735)	
Student Urbanity * Child in Care (CIC) in Child and Family Services	-1.7510	(0.9321)	+
Child in Voluntary Family Services (VFS)	2.6172	(0.5032)	
Student Urbanity * Child in Voluntary Family Services (VFS)	-1.3121	(0.7247)	+
Child in Family Receiving Income Assistance (IA)	2.6894	(0.4349)	***
Student SES (SEFI) * Child in Family Receiving Income Assistance (IA)	-0.4886	(0.2519)	+
Student Urbanity * Child in Family Receiving Income Assistance (IA)	-1.4488	(0.5687)	*
Random Effects (Variance)			
At Level 2 (Student):			
Intercept	9.5080	(0.7344)	***
Child in Voluntary Family Services (VFS)	8.6800	(2.9927)	**
At Level 3 (Class):			
Intercept	2.9364	(0.4527)	***
Child in Voluntary Family Services (VFS)	4.4748	(2.3419)	*
Child in Family Receiving Income Assistance (IA)	3.4380	(1.5071)	*
At Level 1:			
Residual (Intra-personal variance)	33.1671	(0.7488)	***
Fit Statistics			
Deviance	53077.1		
AIC	53123.1		
AICC	53123.3		
BIC	53208.7		

Note: p-value<0.1 +, p-value<0.05 *, p-value<0.01 **, p-value<0.001 ***

Several conclusions are drawn from this model. Students who received PAX training are those who have higher predicted initial SDQ total difficulty scores (higher by 1.6053), but they have bigger improvements by reducing SDQ total scores 0.6235 more than the students in the control group. In addition, female students' anticipated SDQ total scores are 3.4107 points lower at the pre-PAX stage, but male students have a bigger decrease in SDQ scores by 0. 7772 over the course of the study period. Students with one unit higher SEFI values (i.e. coming from families with lower SES) have 1.1113 higher predicted difficulty scores, while these students' SDQ scores will decrease more rapidly over time. Urban students have a lower predicted initial SDQ score by -0.2293; besides, urban students receiving Child in Care services, Voluntary Family Services and Income Assistance have significantly lower expected SDQ scores than the rural students by 1.751, 1.3121 and 1.4488 respectively. Although on average, students receiving Child in Care services, Voluntary Family Services and Income Assistance have generally higher initial SDQ scores (0.7062, 2.6172 and 2.6894). Finally, for students with one unit higher SEFI values (i.e. coming from families with lower SES), their anticipated SDQ scores will be significantly decreased by 0.4886 if their family receives Income Assistance. In terms of the random effects, all systematic variations in the SDQ total scores are significant and evaluated to be 33.1671 overtime for students, 9.5080 among students and 2.9364 for among classes. There also exist significant variations in the SDQ scores of students in terms of different status of student' receipt of Voluntary Family Services (the variances equal to 8.6800 at the inter-student level and 4.4748 at between-class level) and Income Assistance (the variance equal to 3.4380 at the between-class level).

Results in this chapter show that the students with mental disorders tend to have higher SDQ total difficulty scores, and boys are more prone to have elevated scores compared to girls. In terms of cut-offs developed by previous studies, we found that the sensitivity is not up to the desired level for mental disorder screening purposes. Our proposed cut-offs can serve screening purposes better than those by previous studies. We show that adding the SDQ impact scale can improve the prediction of mental disorders compared to only employing the SDQ symptom score. Our results also reveal that the PAX program can decrease the level of mental health problems and the likelihood of being with abnormal mental health. Finally, we observe that several other factors such as gender, living in urbanity, socioeconomic status, and receiving family-supporting services or income assistance are associated with mental health problems.

Chapter 6. Discussion

In this chapter, we discuss our results reported in Chapter 5. The strengths and limitations of this research are also presented. At last, we outline several future directions.

6.1 Study Summary

In this study, we carried out the examination for the applicability of SDQ as a mental disorder screening tool for Canadian children. For this purpose, we integrate data from multiple sources, including a large-scale prevention study (the PAX pilot study), health administrative and family services data. We employed an exhaustive listing approach to find the optimal cut-off based on several criteria. We also showed that the combination of the SDQ components (SDQ domain scores and impact scale) can improve the prediction of mental disorders. Last, we evaluated the PAX program using SDQ.

In terms of the previous-developed cut-off to identify students with mental disorders, the Goodman's and Danish cut-offs are carefully investigated. Goodman's cut-off stands for the classic and the authoritative cut-off developed for the British culture, while the Danish classification bands represent the study result for Demark. Prevalence of mental disorder is determined using both cut-off approaches, the general agreement is achieved with the Canadian Statistics that one in five children in Canada has mental problems. Statistical validation followed after the prevalence summary. Goodman's and Danish cut-off showed relatively high accuracy in classifying Canadian mentally ill students. However, there is an unignorable number of false-negative cases, which is against our research goal to identify as many students with mental disorders as possible. As a result, we ran the exhaustive listing of all possible threshold values and outlined an optimal cut-off to address this issue. Our cut-off threshold showed improvements in identifying more true positive cases without falsely classifying many negative cases. Our cut-off exhibits strength in quick detection of students with potential mental disorders, as the screening procedure is as easy as filling out the questionnaire, adding up the score, and comparing it with our proposed cut-off table.

Moreover, we showed that we can improve the mental disorder prediction with SDQ by adding/using combinations of different SDQ components, including SDQ total difficulty score, SDQ subscale scores, and the impact scale (Q26). Models that contain all five subscales or all 25 itemized questions were anticipated to have the most comprehensive and specific information on the students' mental wellness, however, most of the predictors in these two models are insignificant. Therefore, we selected the approach integrating the total difficulty and impact scale. Additionally, we found that the hyperactivity subscale is excellent to detect students with ADHD, and the detection of this type of mental disorder can be further improved by adding an impact-scale predictor. After summarizing the analytical results from all examined models, we identified more cut-offs based on the predictive model probabilities. These cut-offs are outlined by fitting various logistic models with several predictors adjusted, which was attested to give more classification capacity.

We also discussed the implication of using the SDQ tools for program evaluation by investigating the effectiveness of the PAX program with our proposed cut-off thresholds. To account for the nested data structure, the models were extended to three hierarchies (intra-personal, inter-personal/within-class, and between-class). These mixed effect models first affirmed that using our proposed cut-off for modeling would give us reduced variance components in the higher-level sub-models compared to the previously developed cut-offs. Then, the models validated that the PAX program is effective as the likelihood of students being mentally abnormal is significantly decreased for those who received the PAX program.

Finally, we also explored other factors associated with the risk of being abnormal and who are more likely to have higher mental difficulties. In general, female students have on average lower total difficulties than male students. The students whose families received supporting services such as CIC, VFS and income assistance are likely those having more challenges in mental health determined by SDQ. Children living in rural or poorer socioeconomic areas are more prone to mental difficulties.

All the models used in this thesis are subject to model-specific assumptions. For the multi-level models, we assumed homoscedasticity for the residuals at each level (i.e. the variance of residuals is equal across groups), the independent variance-covariance structure for the random effects, and normal distribution for the model residuals. In addition, for the logistic regression models, linearity between the continuous predictor and the logit of the outcome, and the independence between covariates were assumed. We tested variance assumptions for both model types, and an agreement was achieved for most models. The normality assumption for residuals was not fully satisfied, but the results are, fortunately, acceptable in general due to the large sample size we used and the Centre Limit Theorem for making the sample mean close to a normal distribution.

6.2 Strengths of the Study

This study showed several strengths. First of all, this study used population-representative samples of filled questionnaires accompanied by the mental outcomes and covariate factors extracted from the administrative database for this study cohort. Due to the participation of a large number of schools and students from all regions of Manitoba, there is a reasonable amount of generalizability of our study results to the Manitoba children. Plus, this province-wide study provides a comprehensive understanding of the involving cohort by collecting questionnaire responses at multiple time points: before the PAX program, after the PAX program, and followed up at the Grade Five. Regarding the model examination process, the cut-off thresholds are proposed with appropriate approaches and well-validated; specifically, the one-dimensional cut-off approach was backed up by the Gold Standard validation, while the probability cut-offs were outlined by the logistic models accounted for the non-independent, hierarchical data structure and was validated using training and testing sets. Our cut-off showed high sensitivity without losing an unacceptable amount of specificity. Besides, the minimal cost approach provides a more macro-economical perspective to investigate the cut-offs about how the financial burden incurred by the mental problem can be reduced or even minimized. Together with the innovative outline of the ADHD-targeted detection using SDQ, this research demonstrated an excellent potential to ease the work for teachers and parents to be aware of the children's mental wellness and seek prompt support when needed. Finally, the models give insights into what factors are associated with mental wellness and who should be targeted for mental health intervention.

6.3 Limitations of the Study

Nevertheless, this thesis has limitations. While Kersten et al. (2016) argued that there is only moderate consistency between the SDQ questionnaire responses returned by different informants, only teachers' evaluation of students' mental health was carried out throughout the study. Besides, teachers acted as both the PAX training providers and the evaluators after the program took place, which could bring in the biases towards the overly reported effectiveness of the PAX program. In terms of the Gold Standard used in this research, the mental outcomes/disorders were extracted from the administrative database, and these outcomes could suffer from an underestimation issue of the actual mental disorders in the young population. For instance, it is possible that some of the children with mental difficulties were not present in the clinics/hospitals, prescribed with selective medication for the illness, and/or treated with mental specialists, likely due to the unawareness and/or stigma. However, their mental illness may be reflected by SDQs, which might make the SDQ scores not fully represent the mental issues. This further introduced the possibility that there hardly exists differences in various types of mental disorders like conduct disorder, mood and anxiety disorder, and emotional and behavioral disorder from the linked data. Therefore, only ADHD and the aggregated mental disorder were investigated; no conclusion in other types of mental disorders was drawn from this study. Apart from that, the disease identification is based on the diagnostic codes matched to the health concerns when billing services or prescriptions, which means the person with a diagnostic code does not necessarily have the disease as it is not an actual clinical diagnosis. Moreover, the students with or without mental outcomes are not balanced (i.e. there are much more students without mental disorders compared to those with mental disorders), which brought hardness in making decisions on cut-offs using the classic accuracy measures. Many balancing methods were employed such as undersampling, oversampling, and synthetic sampling using the ROSE and SMOTE packages, and support vector machines using the e1071 package in R, but few improvements showed with these approaches. Finally, we acknowledge that there were students who missed the survey at pre-PAX, post-PAX, and/or the G5 follow-up. These students' might differ from those who completed the study.

6.4 Future Directions

As described in the previous paragraph, there were missing data in the collected responses; hence, it could be complicated but entrancing to employ appropriate missing data imputation before conducting analyses. Anticipated larger study effect size or more significant results and their implication could be obtained with the less amount of information loss. In terms of program evaluation, we modeled the SDQ difficulty scores as both categorical and continuous variables. Considering the data availability of the different time when students could develop mental disorders over the course of the study period, it may be interesting to compare the disease-free/survival period of the students in the PAX group versus those in the control group and recurrence behavior of mental disorders using time-to-event analyses. Or, one could also explore the change from pre-test to post-test and the change from post-test to grade 5 to examine the short-term versus long-term effect of the PAX program separately. Finally, one could also consider the classification method using machine learning models such as decision tree/random forest, k nearest neighbors, and neural network in replace with the logistic regression cut-off approach used in this paper.

Chapter 7. Impact and Significance

It is acknowledged that poor population mental health has negative impacts on the individual life trajectory and government budget. This study bridges the knowledge gap of using the SDQ scoring metrics to capture the potential mental health issues at an early stage or identify the Canadian children that would benefit from mental health assessment and early intervention. As well, the evidence from the PAX effectiveness study can help evaluate mental health prevention programs in Canada and to understand what factors (e.g. gender, living in urbanity, socioeconomic status, and receiving supporting services) might facilitate the implementation of mental health programs and mental health in general.

At the micro level, educators and practitioners can employ SDQ to quickly assess the mental health of children without expertise in the psychiatric field. Then, early intervention and additional supports can be provided to promote healthy development, instead of the mental conditions being left undetected and untreated until the children are in an irreversible mental health state. On the contrary, unnecessary consultation with specialists can be avoided with these scoring metrics. Also, school professionals can pay more attention to those children that are more prone to mental difficulties and provide more targeted and focused intervention. At the macro level, the healthcare planners and policymakers can monitor mental wellness using the SDQ symptom scores. From there, they can decide whether this issue has brought up a big concern and take action on strategy planning accordingly. For instance, the policymakers can discuss when, where and to which group they will implement the mental health promotion programs to improve population mental well-being and reduce health costs. Ultimately, we believe our study facilitates better use of health care resources.

References

American Psychiatric Association. (2018). *What Is Mental Illness?* https://www.psychiatry.org/patients-families/what-is-mental-illness

Arnfred, J., Svendsen, K., Rask, C., Jeppesen, P., Fensbo, L., Houmann, T., Obel, C., Niclasen, J., & Bilenberg, N. (2019). Danish norms for the Strengths and Difficulties Questionnaire. *Danish Medical Journal*, *66*(6), 3–9. http://sdq-dawba.dk/sdq/opgorelse-af-sdq-scoring/

Bland, J. M. (2004). Cluster randomised trials in the medical literature: Two bibliometric surveys. *BMC Medical Research Methodology*, *4*, 1–6. https://doi.org/10.1186/1471-2288-4-21

Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal Impact of Two Universal Preventive Interventions in First Grade on Educational Outcomes in High School. In *Journal of Educational Psychology* (Vol. 101, Issue 4, pp. 926–937). American Psychological Association. https://doi.org/10.1037/a0016586

Brownell, M., Thomson, T., Chartier, M., Towns, D., Au, W., Hong, S., Schultz, J., Scott McCulloch, M., Dale Stevenson, M., Susan Burchill, M., Teresa Mayer, Bm., Jessica Jarmasz, C., & Young, V. (2018). *The PAX Program in Manitoba: A Population-Based Analysis of Children's Outcomes*. http://mchp-appserv.cpe.umanitoba.ca/deliverablesList.html

Canadian Institute for Health Information. (2020). *Child and youth mental health in Canada — Infographic*. https://www.cihi.ca/en/child-and-youth-mental-health-in-canada-infographic

Chartier, M., Brownell, M., Macwilliam, L., Valdivia, J., Nie, Y., Ekuma, O., Burchill, C., Hu, M., Rajotte, L., & Kulbaba, C. (2016). *The Mental Health of Manitoba's Children*.

Claassen, J. A. H. R. (2005). The gold standard: not a golden standard. *BMJ : British Medical Journal*, *330*(7500), 1121. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC557893/

Commonwealth Department of Health and Aged Care. (2000). *Promotion, Prevention and Early Intervention for Mental Health—A Monograph*. Mental Health and Special Programs Branch, Commonwealth Department of Health and Aged Care.

Donovan, C. L., & Spence, S. H. (2000). Prevention of childhood anxiety disorders. In *Clinical Psychology Review*. https://doi.org/10.1016/S0272-7358(99)00040-9

Durlak, J., & Wells, A. (1997). Primary prevention mental health programs for children and adolescents are effective. *American Journal of Community Psychology*, *25*(2), 115–152. https://doi.org/10.1080/09638239817842

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal*. https://doi.org/10.1002/bimj.200410135

Glen, S. (2017). *Gold Standard Test or Procedure*. StatisticsHowTo. https://www.statisticshowto.datasciencecentral.com/gold-standard-test/

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, *9*(2), 129–134.

https://doi.org/10.1007/s007870050008

Goodman, Robert. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *J. Child Psychol. Psychiat*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, Robert. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(5), 791–799. https://doi.org/10.1017/S0021963099004096

Goodman, Robert, Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, *177*(6), 534–539. https://doi.org/10.1192/bjp.177.6.534

Government of Manitoba, & Healthy Child Manitoba Office. (n.d.-a). *PAX Pilot Overview*. Retrieved April 3, 2019, from https://www.gov.mb.ca/healthychild/pax/index.html

Government of Manitoba, & Healthy Child Manitoba Office. (n.d.-b). *The Grade 5 Mental Health Survey*. Retrieved April 3, 2019, from https://www.gov.mb.ca/healthychild/g5mentalhealth/index.html

Government of Manitoba, & Healthy Child Manitoba Office. (n.d.-c). *The Grade 5 Mental Health Survey Report 2015-2016*. Retrieved August 26, 2019, from https://www.gov.mb.ca/healthychild/g5mentalhealth/g5_1516/201516_g5provincialreport.pdf

Hashimoto, R., Brodt, E., Skelly, A., & Dettori, J. (2014). Administrative Database Studies: Goldmine or Goose Chase? *Evidence-Based Spine-Care Journal*, *05*(02), 074–076. https://doi.org/10.1055/s-0034-1390027

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons. https://doi.org/10.1002/9781118548387

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The Distal Impact of Two First-Grade Preventive Interventions on Conduct Problems and Disorder in Early Adolescence. *Journal of Emotional and Behavioral Disorders*, *9*(3), 146–160. https://doi.org/10.1177/106342660100900301

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., Petras, H., Ford, C., Windham, A., & Wilcox, H. C. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence*, *95*, S5–S28. https://doi.org/10.1016/j.drugalcdep.2008.01.004

Kersten, P., Czuba, K., Mcpherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, *40*(1), 64–75. https://doi.org/10.1177/0165025415570647

Kutcher, S., McLuckie, A., & Mental Health Commission of Canada. (2010). *Evergreen: A Child and Youth Mental Health Framework for Canada* (Issue July). https://doi.org/Cat.:

H21-296/2007E-PDF ISBN: 978-0-662-46456-3

Manitoba Centre for Health Policy. (n.d.-a). *About MCHP*. Retrieved October 18, 2019, from http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/about.html

Manitoba Centre for Health Policy. (n.d.-b). *Term: Administrative Health Data*. Retrieved October 18, 2019, from http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=102210

Manitoba Centre for Health Policy. (2002). *Concept: Urban / Rural Definitions*. http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1253

Manitoba Centre for Health Policy. (2020). *Term: Urban or Rural Residence*. http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=104872

Mental Health Commission of Canada. (n.d.-a). *Children and Youth, Overview*. Retrieved March 25, 2019, from https://www.mentalhealthcommission.ca/English/what-we-do/children-and-youth

Mental Health Commission of Canada. (n.d.-b). *Making the Case for Investing in Mental Health in Canada*. Retrieved March 25, 2019, from https://www.mentalhealthcommission.ca/sites/default/files/2016-06/Investing_in_Mental_Health_FINAL_Version_ENG.pdf

Mental Health Commission of Canada. (2015). *Informing the Future : Mental Health Indicators for Canada*. https://www.mentalhealthcommission.ca/sites/default/files/Informing%252520the%252520Future%252520-%252520Mental%252520Health%252520Indicators%252520for%252520Canada_0.pdf

Metge, C., Chateau, D., Prior, H., Soodeen, R.-A., Coster, C. De, & Barre, L. (2009). *Composite Measures/Indices of Health and Health System Performance.* (Issue August).

Neil, A. L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. *Clinical Psychology Review*, *29*(3), 208–215. https://doi.org/10.1016/j.cpr.2009.01.002

Ojemann, R. H., Levitt, E. E., & Lyle Jr, W. H. (1955). The Effects of a “Causal” Teacher-Training Program and Certain Curricular Changes on Grade School Children. *The Journal of Experimental Education*, *24*(2), 95–114. http://www.jstor.org/stable/20153994

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press Oxford.

Perkins, N. J., & Schisterman, E. F. (2005). The Youden index and the optimal cut-point corrected for measurement error. *Biometrical Journal*. https://doi.org/10.1002/bimj.200410133

Perkins, N. J., & Schisterman, E. F. (2006). The Inconsistency of " Optimal " Cut-points Using Two ROC Based Criteria. *Am J Epidemiol*. https://doi.org/10.1093/aje/kwj063

Poduska, J. M., Kellam, S. G., Wang, W., Brown, C. H., Ialongo, N. S., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*, *95*, S29–S44. https://doi.org/10.1016/j.drugalcdep.2007.10.009

Romano, P. S., Peter Hussey, & Ritley, D. (2010). *Selecting Quality and Resource Use Measures: A Decision Guide for Community Quality Collaboratives*. http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/perfmeaspt2.html#q7

Sammut, C., & Webb, G. I. (Eds.). (2011). Accuracy. In *Encyclopedia of Machine Learning* (pp. 9–10). Springer US. https://doi.org/10.1007/978-0-387-30164-8_3

Sarrazin, M. S. V., & Rosenthal, G. E. (2012). Finding Pure and Simple Truths With Administrative Data. *Journal of American Medical Association*, *307*(13), 1433–1435. https://doi.org/doi:10.1001/jama.2012.404

Shojaei, T., Wazana, A., Pitrou, I., & Kovess, V. (2009). The strengths and difficulties questionnaire: validation study in French school-aged children and cross-cultural comparisons. *Social Psychiatry and Psychiatric Epidemiology*, *44*(9), 740–747. https://doi.org/10.1007/s00127-008-0489-8

Silva, T., Osório, F., & Loureiro, S. (2015). SDQ: discriminative validity and diagnostic potential. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00811

Smetanin, P., Stiff, D., Adair, C. E., Ahmad, S., & Khan, M. (2011). The life and economic impact of major mental illnesses in Canada: 2011 to 2041. *RiskAnalytica, on Behalf of the Mental Health Commission of Canada 2011.*, *December*, 199.

Unal, I. (2017). Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Computational and Mathematical Methods in Medicine*, *2017*, 1–14. https://doi.org/10.1155/2017/3762651

Waddell, C. (2017). *Children's Mental Health Policy For the One and the Many Children's Mental Health Policy*. https://www.cmho.org/images/2017 Conference Presentations/Keynotes/Waddell DS.pdf

Washington State Institute for Public Policy. (2018). *Good Behavior Game*. https://www.wsipp.wa.gov/BenefitCost/Program/82

Weare, K., & Nind, M. (2011). Mental health promotion and problem prevention in schools: What does the evidence say? *Health Promotion International*, *26*(SUPPL. 1). https://doi.org/10.1093/heapro/dar075

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., Lui, M., & Tannock, R. (2004). The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child & Adolescent Psychiatry*, *13*(Suppl 2), ii47–ii54. https://doi.org/10.1007/s00787-004-2008-0

World Health Organization. (2002). *Prevention and Promotion in Mental Health*.

Youth Mental Health Canada. (n.d.). *Youth Mental Health Stats in Canada*. Retrieved March 27, 2019, from https://ymhc.ngo/ymh-stats/

Youthinmind. (2015). *Scoring the Strengths & Difficulties Questionnaire for age 4-17*. Youthinmind. https://doi.org/file:///Users/Emily/Downloads/SDQ_English(UK)_4-

17scoring.pdf

YouthinMind. (n.d.-a). *What is the DAWBA?* Retrieved October 7, 2019, from https://youthinmind.com/products-and-services/dawba/

YouthinMind. (n.d.-b). *What is the SDQ?* Retrieved October 7, 2019, from https://youthinmind.com/products-and-services/sdq/

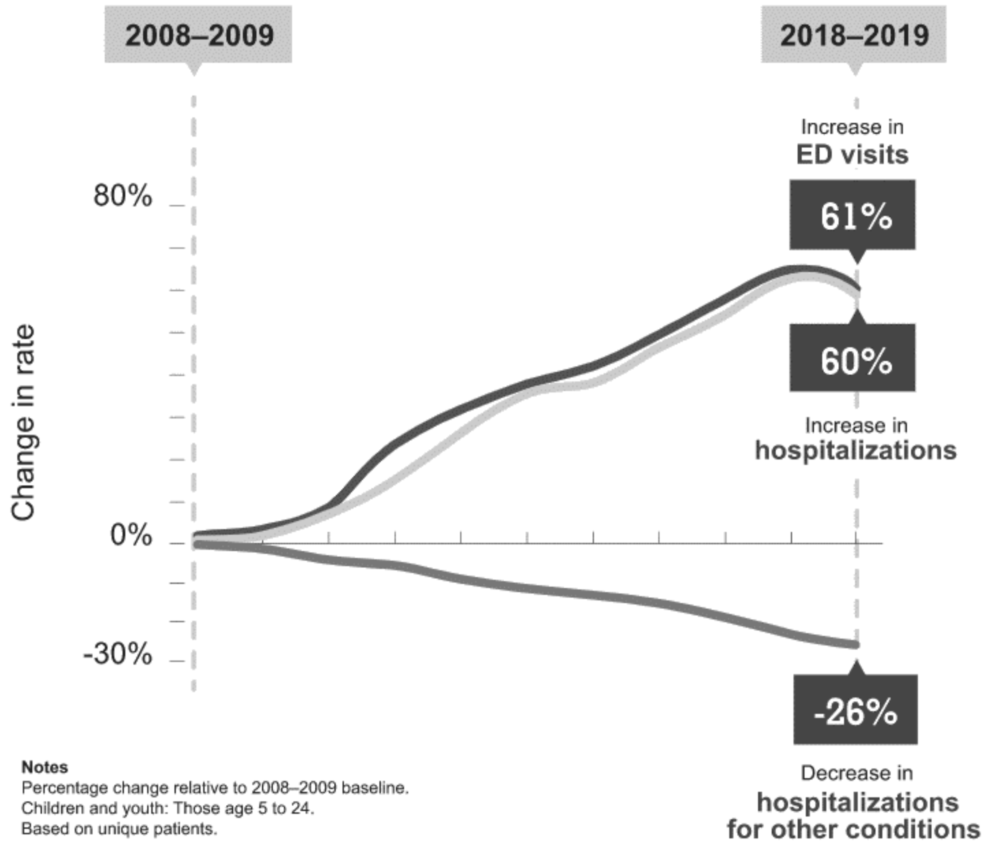
YouthInMind. (n.d.). *Searchable database of SDQ publications - Canada*. Retrieved April 7, 2021, from https://sdqinfo.org/py/sdqinfo/f0.py?t1=&s1=abstract&ao1=or&t2=&s2=title&ao2=or&t3=&s3=title&ao3=or&t4=&s4=title&limitcountry=yes&country=Canada&minyear=1997&maxyear=2018&ageopt=some&minage=1&maxage=18&population=1&population=2&population=3&population=4&p

Yuan, Y., Su, W., & Zhu, M. (2015). Threshold-Free Measures for Assessing the Performance of Medical Screening Tests. *Frontiers in Public Health*, *3*(April), 1–9. https://doi.org/10.3389/fpubh.2015.00057

Zaun, K., Elchehimi, A., Campbell, A., Garron, D., Goodwin, A., Laliberte, P., Lehal, S., Mahleka, D., Robinson, K., Saddleback, J., Sadkowski, M., Savoie, N., & Setter, V. (2016). The Mental Health Strategy For Canada: A Youth Perspective. In *The Mental Health Commission of Canada*. Mental Health Commission of Canada. http://www.mentalhealthcommission.ca/sites/default/files/2016-07/Youth_Strategy_Eng_2016.pdf

Appendix A

Appendix A-1. The Trend of ED Visits and Hospital Admissions Related to Mental Disorder (2006/07-2017/18)



Source: Canadian Institute for Health Information

Appendix B

Appendix B-1. Normative Bandings of **Parent-reported** SDQ Scores for Children from French, UK and US Representative Samples

	Normal			Borderline			Abnormal		
	France[a] (%)	UK[b] (%)	US[c] (%)	France (%)	UK (%)	US (%)	France (%)	UK (%)	US (%)
Emotional symptoms	0–3 (81)	0–3 (81)	0–3 (86)	4 (8)	4 (8)	4 (6)	5–10 (11)	5–10 (11)	5–10 (8)
Conduct problems	0–2 (76)	0–2 (75)	0–2 (81)	3 (12)	3 (12)	3 (8)	4–10 (12)	4–10 (12)	4–10 (11)
Hyperactivity-inattention	0–5 (80)	0–5 (76)	0–5 (84)	6 (8)	6 (8)	6 (5)	7–10 (12)	7–10 (16)	7–10 (11)
Peer relationship problems	0–3 (84)	0–2 (79)	0–2 (79)	4 (9)	3 (10)	3 (10)	5–10 (7)	4–10 (11)	4–10 (11)
Prosocial behaviour	7–10 (84)	8–10 (81)	8–10 (77)	6 (9)	7 (10)	6–7 (11)	0–5 (7)	0–6 (10)	0–5 (13)
Total difficulties	0–13 (82)	0–13 (82)	0–11 (81)	14–16 (10)	14–16 (9)	12–15 (9)	17–40 (8)	17–40 (10)	16–40 (10)
Impact	0 (80)	0 (85)	0 (88)	1 (9)	1 (7)	1 (4)	2–10 (11)	2–10 (8)	2–10 (8)

[a]Regional sample of 1,348 children aged 6–11
[b]National sample of 5,855 children aged 5–10
[c]National sample of 4,843 children aged 4–10

Source: T. Shohei, A. Wazana, I. Pitrou et al.

Appendix B-2. Disease Definition of Child Mental Health Indicators by the Manitoba Centre for Health Policy

Indicator	Definition
Attention-Deficit Hyperactivity Disorder (ADHD)	MCHP has defined ADHD as: • one or more hospitalizations with a diagnosis of hyperkinetic syndrome (ICD-9-CM code 314 or ICD-10-CA code F90) in one fiscal year OR • one or more physician visits with a diagnosis of hyperkinetic syndrome (ICD-9-CM code 314) in one fiscal year OR • two or more prescriptions for ADHD drugs in one fiscal year without a diagnosis in the same fiscal year of: - conduct disorder (ICD-9-CM code 312 or ICD-10-CA codes F63, F91, F92) or - disturbance of emotions (ICD-9-CM code 313 or ICD-10-CA codes F93, F94) or - cataplexy/narcolepsy (ICD-9-CM code 347 or ICD-10-CA code G47.4); OR • one prescription for ADHD drugs in one fiscal year AND a diagnosis of hyperkinetic syndrome in the previous three years. The lists of ADHD medications used in these reports varies: • the ATC code N06BA; or • the generic product name of DEXTROAMPHETAMINE or AMPHETAMINE.

	The definition is restricted to residents ages 3 and older.
Conduct Disorder	MCHP has defined conduct disorder as either: • one or more hospitalizations with a diagnosis of conduct disorder: - ICD-9-CM code 312 or - ICD-10-CA codes All F91 except F91.3; OR • one or more physician visits with a diagnosis of conduct disorder: - ICD-9-CM code 312. The definition is restricted to residents ages 3 and older.
Mood and Anxiety Disorder Mood and Anxiety Disorder Cont'd	A child is considered to have a mood or anxiety disorder if they meet one of the following criteria: • one or more hospitalizations with a diagnosis for depressive disorder, affective psychoses, neurotic depression or adjustment reaction: - ICD-9-CM codes 296.1-296.8, 300.4, 309, 311 or - ICD-10-CA codes F31, F32, F33, F34.1, F38.0, F38.1, F41.2, F43.1, F43.2, F43.8, F53.0, F93.0, F93.1, F93.2; OR with a diagnosis for an anxiety state, phobic disorders or obsessive-compulsive disorders: - ICD-9-CM codes 300.0, 300.2, 300.3, 300.7 or - ICD-10-CA codes F40, F41.0, F41.1, F41.3, F41.8, F41.9, F42, F45.2; • one or more hospitalizations with a diagnosis for anxiety disorders: - ICD-9-CM code 300 or - ICD-10-CA codes F32, F34.1, F40, F41, F42, F44, F45.0, F45.1, F45.2, F48 AND - one or more prescriptions for an antidepressant or mood stabilizer, including medications with the ATC codes N05AN01, N05BA, N06A; • one or more physician visits with a diagnosis for depressive disorder or affective psychoses: - ICD-9-CM codes 296, 311; • one or more physician visits with a diagnosis for anxiety disorders: - ICD-9-CM code 300 AND - one or more prescriptions for an antidepressant or mood stabilizer, including medications with the ATC codes N05AN01, N05BA, N06A; • three or more physician visits with a diagnosis for anxiety disorders or adjustment reaction: ICD-9-CM code 300, 309. The definition is restricted to residents ages 3 and older.

Source: M. Brownell, T. Thomson, M. Chartier et al.

Appendix C

Strengths and Difficulties Questionnaire

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain. Please give your answers on the basis of the child's behavior over the last six months or this school year.

Child's name .... Male/Female

Date of birth...

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children, for example toys, treats, pencils	☐	☐	☐
Often loses temper	☐	☐	☐
Rather solitary, prefers to play alone	☐	☐	☐
Generally well behaved, usually does what adults request	☐	☐	☐
Many worries or often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, depressed or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often offers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets along better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Good attention span, sees work through to the end	☐	☐	☐

Do you have any other comments or concerns?

Overall, do you think that this child has difficulties in any of the following areas: emotions, concentration, behavior or being able to get on with other people?

No	Yes- minor difficulties	Yes- definite difficulties	Yes- severe difficulties
☐	☐	☐	☐

If you have answered "Yes", please answer the following questions about these difficulties:

• How long have these difficulties been present?

Less than a month	1-5 months	6-12 months	Over a year
☐	☐	☐	☐

• Do the difficulties upset or distress the child?

Not at all	Only a little	A medium amount	A great deal
☐	☐	☐	☐

• Do the difficulties interfere with the child's everyday life in the following areas?

	Not at all	Only a little	A medium amount	A great deal
PEER RELATIONSHIPS	☐	☐	☐	☐
CLASSROOM LEARNING	☐	☐	☐	☐

• Do the difficulties put a burden on you or the class as a whole?

Not at all	Only a little	A medium amount	A great deal
☐	☐	☐	☐

Signature .... Date ....